# Vol. I of IV (Appx1-533) Nos. 24-1324, 24-1409

# UNITED STATES COURT OF APPEALS FOR THE FEDERAL CIRCUIT

NATERA, INC.,

Plaintiff-Appellee,

v.

NEOGENOMICS LABORATORIES, INC.,

Defendant-Appellant.

Appeals from the United States District Court for the Middle District of North Carolina, No. 1:23-cv-00629; Hon. Catherine C. Eagles

# NON-CONFIDENTIAL JOINT APPENDIX

DEANNE E. MAYNARD JEFFREY A. LAMKEN ROBERT K. KRY SETH W. LLOYD MORRISON & FOERSTER LLP Lauren M. Weinstein WALTER H HAWES IV 2100 L Street NW, Suite 900 Washington, DC 20037 MOLOLAMKEN LLP Tel.: (202) 887-8740 The Watergate, Suite 500 600 New Hampshire Avenue, N.W. DMaynard@mofo.com Washington, D.C. 20037 DARALYN J. DURIE Tel: (202) 556-2000

Counsel for Appellee Natera, Inc.

jlamken@mololamken.com

(Additional Counsel Listed on Inside Cover)

MARCH 18, 2024

MORRISON & FOERSTER LLP 425 Market Street San Francisco, CA 94105 JOHN FRANKLIN MORROW, JR.

WOMBLE BOND DICKINSON
One West Fourth Street
Winston-Salem, NC 27101

Counsel for NeoGenomics Laboratories, Inc.

KEVIN A. SMITH
TARA SRINIVASAN
QUINN EMANUEL URQUHART &
SULLIVAN, LLP
50 California Street, Floor 22
San Francisco, CA 94111

VICTORIA F. MAROULIS KEVIN P.B. JOHNSON QUINN EMANUEL URQUHART & SULLIVAN, LLP 555 Twin Dolphin Shores, Floor 5 Redwood Shores, CA 94065

SANDRA L. HABERNY QUINN EMANUEL URQUHART & SULLIVAN, LLP 865 South Figueroa Street, Floor 10 Los Angeles, CA 90017

JONATHAN E. BARBEE SARA MARGOLIS MOLOLAMKEN LLP 430 Park Avenue New York, NY 10022

Counsel for Appellee Natera, Inc.

<b>Date</b>	<u>Dkt.</u>	Document VOLUME I	<u>Appx</u>
12/27/2023	169	Order Granting Motion for Preliminary Injunction	Appx1
12/27/2023	171	Preliminary Injunction Order Signed by Chief Judge Catherine C. Eagles	Appx22
		U.S. Patent No. 11,190,035	Appx255
		District Court Docket Sheet	Appx246
7/28/2023	1	Complaint	Appx285
	1-1	U.S. Patent No. 11,530,454	Appx311
		<b>VOLUME II</b>	
	1-21	NeoGenomics 2022 Annual Report	Appx707
	1-30	"NeoGenomics Gets Medicare Coverage for Radar Minimal Residual Disease Assay in Breast Cancer"	Appx882
7/31/2023	5	Natera's Motion for Preliminary Injunction	Appx887
	5-1	Proposed Order	Appx891
7/31/2023	6	Natera's Memorandum of Law In Support of Motion for Preliminary Injunction	Appx894
	6-1	NeoGenomics Gets Medicare Coverage for Radar Minimal Residual Disease Assay in Breast Cancer"	Appx929
7/31/2023	7	Declaration of Anup Malani in Support of Natera's Motion for Preliminary Injunction	Appx2418
	7-15	"Key Takeaways from our MRD panels at the BTIG MedTech, Digital Health, Life Science & Diagnostic Tools Conference"	Appx2844

<u>Date</u>	Dkt.	<b>Document</b>	<u>Appx</u>
	7-19	"NeoGenomics, Inc."	Appx2906
	7-22	"We remain positive on MRD testing across cancer types post our recent KOL call – implications for NTRA, GH, ADPT, NEO, EXAS"	Appx2936
	8-4	"Inivata Expects New Minimal Residual Disease, Recurrence Assay to Support Pharma Trials"	Appx3311
	12-21	"NEO ASCO Booth Tour: NGS and RaDaR Key to the Growing Outlook"	Appx7292
	12-23	"Liquid Biopsy: Going Deeper on MRD in Solid Tumors"	Appx7307
7/31/2023	17	Metzker Declaration In Support of Motion for Preliminary Injunction	Appx7545
	17-3	"Free DNA in the Serum of Cancer Patients and the Effect of Therapy"	Appx7658
	17-6	"Analytical performance and validation of an enhanced TAm-Se circulating tumor DNA sequencing assay"	Appx7688
	17-7	"Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA"	Appx7690
	17-8	"Sensitive detection of ctDNA in early-stage non-small cell lung cancer patients with a personalized sequencing assay"	Appx7725
	17-9	"Circulating Tumor DNA and Late Recurrence in High-Risk Hormone Receptor—Positive, Human Epidermal Growth Factor Receptor 2—Negative Breast Cancer"	Appx7727

<u>Date</u>	<u>Dkt.</u> 17-19	Document "Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence"	<u><b>Appx</b></u> Appx7809
7/31/2023	18	Moshkevich Declaration in Support of Motion for Preliminary Injunction	Appx7922
10/19/2023	107	NeoGenomics' Opposition to Motion for Preliminary Injunction	Appx10475
	108-6	Excerpts from Moshkevich Deposition	Appx10774
	108-7	Excerpts from Malani Deposition	Appx10788
	108-9	Excerpts from Metzker Deposition	Appx10822
	109-3	Natera's First Suppl. Responses & Objections to Neo's Interrogatories	Appx10965
	111-8	Beitsch Letter	Appx11263
10/19/2023	112	Sikri Declaration in Support of Opposition to Motion for Preliminary Injunction	Appx11266
	113-2	Transcript from Goldman Sachs Glocal Healthcare Conference	Appx11533
	113-3	Transcript from Morgan Stanley Global Healthcare Conference	Appx11547
	113-9	Presentation for ASCO 2023 Annual Meeting	Appx11678
	113-10	"Residual ctDNA after treatment predicts early relapse in patients with early-stage non-small cell lung cancer"	Appx11694
10/19/2023	115-12	Beitsch Letter	Appx11859
10/19/2023	116	Van Ness Declaration in Support of Opposition to Motion for Preliminary Injunction	Appx12001

<u>Date</u>	<u>Dkt.</u>	Document VOLUME III	<u>Appx</u>
10/19/2023	116	Van Ness Declaration in Support of Opposition to Motion for Preliminary Injunction (cont.)	Appx12136
	116-6	Excerpts from Metzker Deposition	Appx12399
	116-10	Natera's <i>CareDx</i> Opening Brief in Support of Summary Judgment	Appx12582
	116-12	Metzker Illumina Declaration	Appx12958
	116-13	Fluidigm U.S. Patent Application 2010/0120038 A1	Appx13077
	116-14	Kaper Abstract	Appx13109
	116-15	Kaper Slide	Appx13112
	116-17	Cantor Application	Appx13141
	116-18	Ehrich Application	Appx13197
	116-19	Lo Application	Appx13253
	116-20	Quackenbush <i>CareDx</i> Declaration	Appx13369
	119-1	"Oncologist Survey: MRD in Early Stages of Adoption"	Appx16140
11/6/2023	141	Metzker Reply Declaration	Appx18741
	141-2	"Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies"	Appx19065
11/6/2023	144	Natera's Reply in Support of Motion for Preliminary Injunction	Appx19186
	144-5	Malackowski Deposition	Appx19502
11/18/2023	154	Tutorial Hearing Transcript VOLUME IV	Appx19868
11/18/2023	154	Tutorial Hearing Transcript (cont.)	Appx19917
	162-1	Metzker Deposition	Appx20005
	162-2	Tutorial Hearing Transcript	Appx20073

<u>Date</u>	<u>Dkt.</u>	<b>Document</b>	<b>Appx</b>
	162-3	Preliminary Injunction Hearing Transcript	Appx20153
12/8/2023	165	NeoGenomics' Response to Suggestion of Subsequently Decided Authority	Appx20707
	165-1	Notice of Subsequent Development in <i>Natera, Inc. v. ArcherDX</i>	Appx20713
12/27/2023	173	NeoGenomics' Notice of Appeal	Appx20748
12/29/2023	176	NeoGenomics' Motion to Stay Preliminary Injunction	Appx20751
12/29/2023	177	Memorandum of Law in Support of Motion to Stay Preliminary Injunction	Appx20756
12/29/2023	178	NeoGenomics' Motion to Modify or Clarify Preliminary Injunction	Appx20783
12/29/2023	183	Memorandum of Law in Support of Motion to Modify or Clarify Preliminary Injunction	Appx20791
	183-1	Sikri Declaration	Appx20804
1/4/2024	188	Natera's Memorandum of Law in Opposition to Motion to Modify or Clarify Preliminary Injunction	Appx20817
	188-2	"Our thoughts on the preliminary injunction decision in favor of NTRA vs. NEO"	Appx20853
1/8/2024	192	Natera's Memorandum of Law in Opposition to Stay Motion	Appx20884
1/10/2024	193	Order on Motion to Modify/Clarify Preliminary Injunction	Appx20911
1/11/2024	196	NeoGenomics' Reply in Support of Motion to Stay	Appx20918
1/16/2024	197	Natera's Notice of Posting Security	Appx20937
1/26/2024	201	NeoGenomics' Amended Notice of Appeal	Appx20942

<b>Date</b>	Dkt.	<b>Document</b>	<b>Appx</b>
2/2/2024	206	Order on Motion to Modify/Clarify Preliminary Injunction	Appx20945
		NeoGenomics' Technology Tutorial Presentation	Appx20949
	246-1	Natera's Technology Tutorial Presentation	Appx21013
		NeoGenomics' Preliminary Injunction Argument Presentation	Appx21068
2/14/24	216	Order on Motion to Stay Pending Appeal	Appx21320
2/14/24	217	Order on Motion to Modify Preliminary Injunction	Appx21327
8/4/23	49	Scheduling Conference Transcript	Appx21332
		Natera's Preliminary Injunction Argument Presentation	Appx21358
1/31/24	204	NeoGenomics' Proof of Service of Notice of Preliminary Injunction	Appx21520
	204-1	Notice of Preliminary Injunction to Employees	Appx21523
	204-2	Notice of Preliminary Injunction to Sales Force	Appx21527
	204-3	Notice of Preliminary Injunction to External Clients	Appx21531
	204-4	Notice of Copy of Preliminary Injunction to Employees	Appx21533

NOTE: Documents filed under seal in the district court used both green and yellow highlighting to indicate confidentiality. But the yellow highlighting at Appx7624, Appx7926-7929, Appx10490-10492, and Appx12169-12170 has been unsealed in the district court and does not indicate confidential information; at Appx7624 and Appx7926-7929 only the green highlighting indicates confidentiality. The slides at Appx21068-21319 also includes non-confidential markings; only the full-page yellow rectangular markings indicate relevant confidentiality. The redaction appearing at Appx20714 was part of the underlying document as originally filed in the district court. Appx12169 of the non-confidential addendum contains a paragraph number that is highlighted but not redacted; the highlighting does not indicate confidentiality.

# **CONFIDENTIAL MATERIAL OMITTED**

The non-confidential version of this appendix redacts material filed under seal pursuant to the protective order issued by the district court. As required by Federal Circuit Rule 25.1(e)(1)(B), the table below notes the specific pages with redacted material in the non-confidential appendix and the general nature of that material.

**Description of Redacted Material in Non-Confidential Appendix** 

Document	Pages	Description
Metzker Declaration (Dkt. 17)	Appx7624	Natera's confidential product information
Moshkevich Declaration (Dkt. 18)	Appx7926-7929	Natera's confidential product information
NeoGenomics' Response in Opposition to Motion for Preliminary Injunction (Dkt. 107)	Appx10480- 104481, Appx10485, Appx10487, Appx10494-10495, Appx10497-10500, Appx10502-10504	NeoGenomics' confidential financial, technical, and customer information
Excerpts from Moshkevich Deposition (Dkt. 108-6)	Appx10774-10787	The material omitted from page Appx10777 includes information relating to Natera's internal business structure. The material omitted from pages Appx10778-79 includes information relating to Natera's internal market analyses and marketing strategy. The material omitted from pages Appx10780-83 and Appx10785 includes information relating to Natera's business development efforts and customer relationships. The material omitted from pages Appx10786-87 includes information relating to Natera's IP licensing disputes and business relationships.

Sikri Declaration	Appx11274-11275,	NeoGenomics' confidential financial,
(Dkt. 94)	Appx11277-11284,	technical, and customer information
	Appx11286-11294	
Van Ness	Appx12171	Natera's confidential business
Declaration		relationships
(Dkt. 97)		
Metzker	Appx18746	NeoGenomics' confidential technical
Declaration	Appx18898-18901	information
(Dkt. 141)		
Natera's Reply	Appx19187,	The material omitted at Appx19187,
in Support of	Appx19191,	Appx19191, and Appx19194-19195
Motion for	Appx19194-19195,	includes NeoGenomics' confidential
Preliminary	Appx19198,	product information. The material
Injunction (Dkt.	Appx19200-19204	omitted from page Appx19198 includes
144)		a summary of a confidential
		communication regarding inventorship
		of the '454 patent. The material omitted
		from pages Appx19200-19204 includes
		information relating to Natera's
		customer relationships and
		NeoGenomics' clinical research.
Deposition of	Appx19518	The material omitted from page
Malackowski		Appx19518 includes information
(Dkt. 144-5)		relating to NeoGenomics' market
		opportunities.
Sikri Declaration	Appx20805-20806	NeoGenomics' confidential client
(Dkt. 183-1)	Appx20809	financial and customer information
NeoGenomics'	Appx20933	NeoGenomics' confidential financial
Reply in Support		information
of Motion to		
Stay Preliminary		
Injunction		
(Dkt. 196)		
NeoGenomics'	Appx21251	Natera's confidential technical and
Preliminary	Appx21269-21296	business information
Injunction	Appx21313	
Presentation	Appx21316-21318	

ny-2702558

# IN THE UNITED STATES DISTRICT COURT FOR THE MIDDLE DISTRICT OF NORTH CAROLINA

NATERA, INC.,		)	
		)	
	Plaintiff,	)	
		)	1:23-CV-629
V.		)	
		)	
NEOGENOMICS		)	
LABORATORIES, INC.,		)	
		)	
	Defendant.	)	

## **ORDER**

Natera, Inc. and NeoGenomics Laboratories, Inc. are research-focused healthcare companies that operate in the oncology testing industry. Both companies have products that can be used for earlier detection of cancer relapse. Natera seeks a preliminary injunction, contending that NeoGenomics' product, RaDaR, infringes two of Natera's patents. Because Natera has shown a likelihood of success on the merits that NeoGenomics is infringing the '035 patent and the other requirements for injunctive relief are met, the motion will be granted.

For purposes of this Order, the Court makes the following findings of fact and conclusions of law.

#### I. FACTS

As relevant here, Natera holds two method patents. Patent No. 11,519,035 issued on December 6, 2022, Doc. 1-2 at 2, and Patent No. 11,530,454 issued on December 20, 2022. Doc. 1-1 at 2. In simplified terms, the '035 patent provides methods and compositions for amplifying targeted genetic material while reducing amplification of

non-targeted genetic material. *See* Doc. 1-2 at 2, 89. The '454 patent provides methods, systems, and computer readable medium for detecting variations in genetic material indicative of disease or disease recurrence. *See* Doc. 1-1 at 2, 137.

Natera uses the methods described in these two patents in a product marketed under the brand name Signatera. *See* Doc. 9-18 at 2–3.<sup>1</sup> NeoGenomics offers a competing product under the brand name RaDaR. Doc. 94 at ¶ 10. RaDaR has been used in clinical cancer research since April 2020.<sup>2</sup> *Id.* at ¶ 11. It has been commercially available since March 2023. *Id.* at ¶ 15.

Signatera and RaDaR each work by identifying circulating DNA fragments from cancer cells within the bloodstream. *See* Doc. 13 at ¶ 125; Doc. 94 at ¶ 10. The presence of these tumor DNA fragments can indicate the efficacy of cancer treatment and the risk of cancer recurrence. Doc. 7 at ¶ 23; Doc. 13 at ¶ 34.

During a cell's life cycle, it naturally sheds short fragments of DNA into the bloodstream. Doc. 13 at ¶ 36. These DNA fragments are referred to as cell-free DNA (cfDNA). *Id.*; Doc. 97 at ¶ 43. Both healthy cells and cancerous cells create cfDNA. Doc. 13 at ¶ 36. The subset of cfDNA that comes from cancer cells is referred to as circulating tumor DNA (ctDNA). *Id.*; Doc. 97 at ¶ 247.

<sup>&</sup>lt;sup>1</sup> Throughout this Order, the Court has cited some of the evidence in the record that supports its factual findings but has made no attempt to cite all the evidence supporting its findings.

<sup>&</sup>lt;sup>2</sup> RaDaR was initially offered by a company known as Inivata. Doc. 94 at ¶ 11. NeoGenomics acquired Inivata in 2021. *Id.* at ¶ 12.

If a patient has a positive response to cancer treatment, the patient's tumor typically decreases in size, eventually becoming undetectable in radiographic imaging or clinical examination. Doc 13 at ¶ 34. If tumor cells remain in a patient's body after treatment, there is the potential for cancer relapse, either locally or through metastases.

Id. Molecular residual disease (MRD) refers to the presence of small amounts of tumor DNA molecules in the body after treatment. Id.; Doc. 7-38 at 2. Early detection of MRD supports better patient outcomes. Doc. 13 at ¶ 34.

MRD tests are either tumor informed or tumor naïve. Doc. 7 at ¶ 24. Tumor informed tests are designed from a patient's genetic information obtained from a tissue biopsy of the patient's tumor. *Id.* These bespoke tests are often preferred by doctors and are seen as highly sensitive because they are personalized to the patient. *Id.* at ¶¶ 34–35; Doc. 92-1 at 4, 17. Tumor naïve tests can provide faster results but are less favored by doctors because they are perceived to be less accurate. Doc. 7 at ¶¶ 34–35.

Signatera, Natera's product, and RaDaR, NeoGenomics' product, are tumor informed MRD tests that work by detecting trace amounts of ctDNA in a patient's bloodstream. *See* Doc. 13 at ¶¶ 115, 125; Doc. 94 at ¶ 10. Patients known to have cancer first provide tumor tissue samples. Doc. 13 at ¶ 118; Doc. 94 at ¶ 10. Next, the DNA from those tissue samples is sequenced, and the DNA information is used to design liquid biopsy MRD assays. Doc. 13 at ¶¶ 118–19; Doc. 94 at ¶¶ 9–10. After cancer treatment ends, patients regularly provide blood samples for testing. Doc. 13 at ¶ 120; Doc. 94 at ¶ 10. Using tumor informed MRD assays, doctors and scientists can detect DNA from cancer cells in the blood in the form of ctDNA. Doc. 13 at ¶¶ 120–22; Doc. 94 at ¶ 10.

Natera is the leader in the MRD assay market. *See* Doc. 7 at ¶¶ 45, 73; Doc. 92-1 at 2; Doc. 11-9 at 42. According to one report, Natera has 74% of the total market share for both tumor informed and tumor naïve tests. Doc. 92-1 at 12, 17. RaDaR is the only other tumor informed MRD test available for clinical use and covered by private insurance.<sup>3</sup> *See* Doc. 7 at ¶ 48. Other companies, such as Guardant Health, compete in the MRD space, but they offer tumor naïve products, not tumor informed products. *See id.* at ¶ 53; Doc. 92-1 at 17. The MRD testing market is expected to grow substantially over the next few years. Doc. 92-1 at 3–4.

Additional findings of fact will be stated as they become relevant.

#### II. PRELIMINARY INJUNCTION STANDARD

Courts have the power to grant preliminary injunctions in patent infringement lawsuits. See High Tech Med. Instrumentation, Inc. v. New Image Indus. Inc., 49 F.3d 1551, 1554 (Fed. Cir. 1995); 35 U.S.C. § 283. To establish a preliminary injunction is warranted, the patentee seeking an injunction must show: "(1) it is likely to succeed on the merits, (2) it is likely to suffer irreparable harm in the absence of preliminary relief, (3) the balance of equities tips in its favor, and (4) an injunction is in the public interest." BlephEx, LLC v. Myco Indus., Inc., 24 F.4th 1391, 1398 (Fed. Cir. 2022) (quoting Winter v. Nat. Res. Def. Council, Inc., 555 U.S. 7, 20 (2008)) (cleaned up).

<sup>&</sup>lt;sup>3</sup> Until recently, Invitae offered another tumor informed MRD test under the brand name PCM. After a jury in the District of Delaware found that product infringed other patents held by Natera, the court permanently enjoined Invitae from offering PCM for clinical use. *See* Doc. 164-1 (redacted permanent injunction entered in *Natera*, *Inc.* v. *ArcherDX*, *Inc.*, No. 20-CV-125 (D. Del. Nov. 21, 2023)).

#### III. SUCCESS ON THE MERITS

To show a likelihood of success on the merits, the moving party must demonstrate "(1) it will likely prove infringement and (2) its infringement claim will likely withstand challenges to the validity and enforceability of the patents." *See Purdue Pharma L.P. v. Boehringer Ingelheim GMBH*, 237 F.3d 1359, 1363 (Fed. Cir. 2001) (cleaned up). In patent cases, the likelihood of success factor is governed by Federal Circuit law. *See ABC Corp. I v. P'ship and Unincorporated Ass'ns Identified on Schedule "A"*, 52 F.4th 934, 941 (Fed. Cir. 2022). When evaluating the likelihood of success, courts consider all the burdens and presumptions that would apply at trial. *See Purdue*, 237 F.3d at 1363.

To assess the likelihood of infringement, courts first "determine the scope and meaning of the patent claims asserted." *Oakley, Inc. v. Sunglass Hut. Int'l*, 316 F.3d 1331, 1339 (Fed. Cir. 2003) (cleaned up); *see also CommScope Techs. LLC v. Dali Wireless Inc.*, 10 F.4th 1289, 1295 (Fed. Cir. 2021). Then "the properly construed claims are compared to the allegedly infringing device." *Oakley*, 316 F.3d at 1339 (cleaned up). Infringement claims must also withstand any challenges to a patent's validity and enforceability. *See Purdue*, 237 F.3d at 1363.

# IV. '035 PATENT – LIKELIHOOD OF SUCCESS ON THE MERITS A. Infringement

For purposes of the preliminary injunction motion, Natera asserts that the RaDaR product offered by NeoGenomics infringes claims 1, 12, and 13 of the '035 patent. Doc. 71-4 at 2. Claim 1 states:

A method for amplifying and sequencing DNA, comprising: tagging isolated cell free DNA with one or more universal tail adaptors to generate tagged products, wherein the isolated cell-free DNA is isolated from a blood sample collected from a subject who is not a pregnant woman;

amplifying the tagged products one or more times to generate final amplification products, wherein one of the amplification steps comprises targeted amplification of a plurality of single nucleotide polymorphism (SNP) loci in a single reaction volume, wherein one of the amplifying steps introduces a barcode and one or more sequencing tags; and

sequencing the plurality of SNP loci on the cell free DNA by conducting massively parallel sequencing on the final amplification products, wherein the plurality of SNP loci comprises 25–2,000 loci associated with cancer.

Doc. 1-2 at 213. Amplification refers to increasing "the number of copies of a molecule, such as a molecule of DNA." *Id.* at 109.

Natera has made a strong showing that the RaDaR test made and sold by NeoGenomics uses the method claimed in the '035 patent and infringes the '035 patent. It is likely to succeed on the merits of its infringement claim.

NeoGenomics contends that the claims at issue require targeted amplification of already tagged DNA and that RaDaR does not use targeted amplification on tagged products. *See* Doc. 89 at 11–12; Doc. 97 at ¶¶ 89–93. But this ignores the fact that RaDaR first tags the products with the CS1 adaptor sequence, then performs targeted amplification to tag the products a second time with the CS2 sequence. *See* Doc. 13 at ¶¶ 44, 88 (explaining RaDaR uses teachings of Forshew); Doc. 145-1 at 57, 60 (discussing PCR amplification and CS1 and CS2 tagging in Forshew). Thus, RaDaR amplifies "the tagged products one or more times to generate final amplification products, wherein one of the amplification steps comprises targeted amplification." Doc.

1-2 at 213. Natera has shown a likelihood of success on the merits on its claim that NeoGenomics infringes Claim 1 of the '035 patent.

NeoGenomics also contends that the dependent Claims 12 and 13 require that the tagging referenced in Claim 1 occur over two rounds of polymerase chain reaction (PCR)<sup>4</sup> because Claims 12 and 13 refer to a first primer comprising a first universal tail adaptor and a second primer comprising a second universal tail adaptor. *See* Doc. 89 at 12; Doc. 1-2 at 213. But an independent claim like Claim 1 is "broader than the claims that depend from it," *Littlefuse, Inc. v. Mersen USA EP Corp.*, 29 F.4th 1376, 1380 (Fed. Cir. 2022), and "each claim in a patent is presumptively different in scope." *Trs. of Columbia Univ. in N.Y. v. Symantec Corp.*, 811 F.3d 1359, 1370 (Fed. Cir. 2016).

When a limitation found in a dependent claim is the only meaningful difference between the dependent and independent claim, "the independent claim is not restricted by the added limitation in the dependent claim." *Id.*; *see also Acumed LLC v. Stryker Corp.*, 483 F.3d 800, 806 (Fed. Cir. 2007) (holding that proposed claim readings should not make the independent and dependent claims identical in scope). This argument does not weaken Natera's likelihood of success on the merits as to infringement of Claim 1.

#### **B.** Validity

To show a likelihood of success, the patent holder must demonstrate a patent is likely to withstand any challenges to validity. *See Purdue*, 237 F.3d at 1363. A patent is presumed valid. *Id.* at 1365; *KSR Int'l Co. v. Teleflex, Inc.*, 550 U.S. 398, 412 (2007).

<sup>&</sup>lt;sup>4</sup> PCR is a common method of amplification. See Doc. 141-6 at 7.

The party challenging the validity of a patent must come forward with evidence that raises a substantial question of validity. *Titan Tire Corp. v. Case New Holland, Inc.*, 566 F.3d 1372, 1376 (Fed. Cir. 2009). The patentee then has the "burden of responding with contrary evidence." *Id.* at 1377. If the patentee does not prove the validity question "lacks substantial merit," courts will deny the motion for a preliminary injunction. *Amazon.com, Inc. v. Barnesandnoble.com, Inc.*, 239 F.3d 1343, 1350–51 (Fed. Cir. 2001).

NeoGenomics challenges the validity of the '035 patent on several grounds, primarily contending that it is obvious. *See* Doc. 89 at 15–16. NeoGenomics also makes shorthand or conclusory arguments that there are issues with the patent's written description, that Natera did not properly explain changes made to the named inventors, and that the '035 patent covers patent-ineligible subject matter. *See id.* at 16–17, 19.

# 1. Obviousness

A patent cannot issue when "the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious . . . to a person having ordinary skill in the art." 35 U.S.C. § 103; see also Teleflex, 550 U.S. at 406–07. A party seeking to show a patent is invalid on obviousness grounds must show that "a person of ordinary skill in the art would have been motivated to combine or modify the teachings in the prior art and would have had a reasonable expectation of success in doing so." Regents of Univ. of Cal. v. Broad. Inst., 903 F.3d 1286, 1291 (Fed. Cir. 2018).

An invention is not automatically obvious just because a motivation to combine may exist, see Arctic Cat Inc. v. Bombardier Recreational Prods. Inc., 876 F.3d 1350,

1359–60 (Fed. Cir. 2017), or "because all of the claimed limitations were known in the prior art at the time of the invention." *Forest Lab'ys, LLC v. Sigmapharm Lab'ys, LLC*, 918 F.3d 928, 934 (Fed. Cir. 2019). A challenger asserting obviousness must give "some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness." *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006). Mere conclusory statements tend to show hindsight bias, not that the invention is obvious. *ActiveVideo Networks, Inc. v. Verizon Commc'ns, Inc.*, 694 F.3d 1312, 1327 (Fed. Cir. 2012).

The priority date for the claims of the '035 patent is May 18, 2011, *see* Doc. 1-2 at 2; Doc. 138-2 at 7–8, so the prior art must be from before this date. NeoGenomics first asserts that the Fluidigm Access Array System, discussed in the 2010 Kaper publication renders the '035 patent obvious. *See* Doc. 89 at 15–16; Doc. 97-14 (abstract); Doc. 97-15 (poster). But Dr. Kaper used DNA samples from tumor tissue, not cfDNA, in discussing Access Array. *See* Doc. 141 at ¶ 156.

NeoGenomics contends that it would have been obvious to modify Access Array for cfDNA because cfDNA was known at this time to be useful for cancer detection. *See* Doc. 89 at 15; Doc. 97 at ¶¶ 240–41. But there were many well-known barriers to using cfDNA. *See* Doc. 141 at ¶ 157 (Tumor tissue samples provide tumor-associated DNA in greater quantity than the ctDNA found in cfDNA.); Doc. 141-2 at 2 (Cell-free DNA is fragmented and exists in low yield within the body.); Doc. 13-7 at 2 (Circulating tumor DNA is just a subset of cfDNA and exists in even lower yields.); Doc 141 at ¶ 16 (Even if a cfDNA molecule has a SNP of interest, the cfDNA may be fragmented such that it "does not contain sites for both primers of a primer pair to bind" and the SNP will not

amplify "successfully when PCR is performed."). These challenges associated with cfDNA, and others, presented obstacles to successfully amplifying and sequencing ctDNA with precision during the relevant time period, *see* Doc. 141-2 (published October 2016); Doc. 141 at ¶¶ 19–21, making it unlikely a person skilled in the art would have been motivated to use cfDNA with Access Array and would have anticipated success in doing so. *See Forest Lab'ys*, 918 F.3d at 934; *Arctic Cat*, 876 F.3d at 1359–60.<sup>5</sup>

NeoGenomics' assertions about Access Array appear to show hindsight bias more than they support a substantial question of obviousness. *ActiveVideo*, 694 F.3d at 1327.

In passing, NeoGenomics also contends that something called "ARM-PCR" renders the '035 patent obvious for "similar reasons." Doc. 89 at 15. This conclusory assertion does not raise a substantial question of validity.

# 2. Written Description

Title 35 U.S.C. § 112 requires a patent specification to contain a written description of the invention that "discloses and teaches" what is claimed. *Ariad Pharms.*, *Inc. v. Eli Lilly and Co.*, 598 F.3d 1336, 1347 (Fed. Cir. 2010); 35 U.S.C. § 112(a). There is no required "particular form of disclosure," and the written description need not use the

<sup>&</sup>lt;sup>5</sup> At oral argument, Natera's counsel said that Natera had also invented ways to overcome these barriers.

<sup>&</sup>lt;sup>6</sup> In support of this half-a-sentence argument, NeoGenomics provides a bulk citation to some 68 paragraphs covering well over 10 pages in a declaration from one of its experts, *see* Doc. 89 at 15, citing Doc. 97 at ¶¶ 235–303. Expert declarations are not substitutes for briefs and implicit attempts to incorporate them by reference cannot be used to avoid the word limits for briefing. *See* discussion *infra* at 13.

exact words of the claims. See Ariad, 598 F.3d at 1352; Univ. of Rochester v. G.D. Searle & Co., Inc., 358 F.3d 916, 922–23 (Fed. Cir. 2004).

NeoGenomics contends that the specification for the '035 patent "lack[s] examples of the claimed processes" and "[t]he claims are disassociated from the described invention." *See* Doc. 89 at 17. This argument, made in one paragraph, does not raise a substantial question of validity.<sup>7</sup>

# 3. Inventorship

The named inventors in an issued patent are presumed correct. *Eli Lilly and Co. v. Aradigm Corp.*, 376 F.3d 1352, 1358 (Fed. Cir. 2004). A "party seeking correction of inventorship must show by clear and convincing evidence that a joint inventor should have been listed." *Blue Gentian, LLC v. Tristar Prods., Inc.*, 70 F.4th 1351, 1357 (Fed. Cir. 2023) (citing *Aradigm*, 376 F.3d at 1358). While a patent must reflect true inventorship, "a patent cannot be invalidated if inventorship can be corrected instead." *Egenera, Inc. v. Cisco Sys., Inc.*, 972 F.3d 1367, 1376 (Fed. Cir. 2020).

Correcting inventorship for an issued patent can be done either by petition to the Director of the U.S. Patent and Trademark Office or by court order. *See* 35 U.S.C.

<sup>&</sup>lt;sup>7</sup> In support of its perfunctory "disassociation" argument, Natera again provides a bulk citation to over 20 paragraphs of an expert's declaration. *See* Doc. 89 at 17, citing Doc. 97 at ¶¶ 404–26. But most of those paragraphs are addressed to the '454 patent, and such bulk citations do not direct the Court's attention with any specificity to facts relevant to the '035 patent. *See* Doc. 27 at ¶ 1 (Order requiring that citations to a multi-page exhibit "must contain a pin cite to the specific page" or "paragraph number"). The Court is not required to sift through blocks of evidence itself to locate the truffles. *See Hughes v. B/E Aerospace, Inc.*, No. 12-CV-717, 2014 WL 906220, at \*1 n.1 (M.D.N.C. Mar. 7, 2014) ("A party should not expect a court to do the work that it elected not to do."). As the Court has stated, *see supra* note 6, at 10 and discussion *infra* at 13, expert declarations are not mechanisms to avoid word limits.

§§ 256(a), (b). Inventorship in a patent application can also be corrected, 35 U.S.C. § 116(c), and the patent office does not require an explanation for the correction. *See* 37 C.F.R. § 1.48 (setting out requirements for correction of inventorship in provisional and nonprovisional patent applications).

NeoGenomics does not show by clear and convincing evidence that an inventor is missing from the '035 patent. *See Blue Gentian*, 70 F.4th at 1357. In fact, NeoGenomics does not raise any issues with any specific inventor of the '035 patent, saying only that the changes Natera made during the patent application process "appear dubious." *See* Doc. 89 at 19. This perfunctory assertion does not raise a substantial question of invalidity for the '035 patent based on inventorship.

# 4. Patent Ineligible Subject Matter

In arguing that the '035 patent is not valid because of ineligible subject matter,
NeoGenomics purports to incorporate by reference an argument made in another brief.
Doc. 89 at 19 (referencing NeoGenomics' motion to dismiss brief, Doc. 52). The Court will disregard this argument for purposes of this motion.

The Local Rules limit briefs in support of motions and responsive briefs to no more than 6,250 words. *See* LR 7.3(d). Per this Court's standard order, when a party incorporates arguments made in other briefs, the incorporating brief's word count is "correspondingly decreased." Doc. 27 at ¶ 1. The word count in the incorporated brief, Doc. 52, is some 6,233 words. The word count in the incorporating brief, Doc. 89, excluding the incorporated argument, is some 6,218 words. NeoGenomics has thus attempted to surpass the word limit by thousands of words.

Word or page limits are common across courts of all stripes for obvious reasons; among other things, such limits require parties to avoid rhetoric and to present their best arguments, and they facilitate prompt, efficient resolution of cases. Efforts to circumvent word limits by incorporating briefs by reference or dividing one motion into several motions are improper. *See, e.g., Basulto v. Netflix, Inc.*, No. 22-CV-21796, 2022 WL 17532279, at \*2 (S.D. Fla. Dec. 8, 2022) (collecting cases); *Monec Holding AG v. Motorola Mobility, Inc.*, No. 11-CV-798, 2014 WL 4402825, at \*2 (D. Del. Sept. 5, 2014) (finding opposing party prejudiced if forced to respond to legal arguments in incorporated exhibit); *Aircraft Tech. Publishers v. Avantext, Inc.*, No. 7-CV-4154, 2009 WL 3833573, \*1 (N.D. Cal. Nov. 16, 2009) (striking summary judgment motions that were improperly filed separately to avoid page limits).

At the preliminary injunction hearing,<sup>8</sup> NeoGenomics focused on the arguments it made in its preliminary injunction opposition brief, not the incorporated brief about patent ineligible subject matter. The Court declines to consider the subject matter arguments in connection with the preliminary injunction motion.

### C. Conclusion

Natera has demonstrated a likelihood of success on the merits for its '035 patent infringement claim. RaDaR likely infringes Claim 1 of the '035 patent, and NeoGenomics has not raised a substantial question of validity.

<sup>&</sup>lt;sup>8</sup> At the Court's request, the deputy clerk had advised the parties several days before the hearing of a number of matters the Court wanted addressed, one of which was the effort to avoid the word limits by incorporating another brief.

#### V. '454 PATENT – LIKELIHOOD OF SUCCESS ON THE MERITS

In view of its findings and conclusions on the '035 patent, the Court need not address whether Natera has shown a likelihood of success on its claim that the '454 patent is valid and infringed.

#### VI. IRREPARABLE HARM

#### A. Likelihood of Irreparable Harm

The plaintiff seeking preliminary injunctive relief must show a likelihood of irreparable harm. *BlephEx*, 24 F.4th at 1398. A patentee suffers irreparable harm when "forced to compete against products that incorporate and infringe its own patented inventions." *Douglas Dynamics*, *LLC v. Buyers Prods. Co.*, 717 F.3d 1336, 1345 (Fed. Cir. 2013). Natera will likely suffer irreparable harm if NeoGenomics continues to offer RaDaR in the marketplace.

Evidence of head-to-head competition, lost market share, lost sales, and a decline in reputation and brand distinction can support a showing of irreparable harm. *See TEK Glob., S.R.L. v. Sealant Sys. Int'l, Inc.*, 920 F.3d 777, 793 (Fed. Cir. 2019); *Douglas Dynamics*, 717 F.3d at 1344. When two competitors directly compete "for the same customers in the same markets," irreparable harm is evident. *See Presidio Components Inc. v. Am. Tech. Ceramics Corp.*, 702 F.3d 1351, 1363 (Fed. Cir. 2012).

Natera and NeoGenomics are direct competitors in the tumor informed MRD marketplace. *See* Doc. 7 at ¶ 131; Doc. 92-1 at 26. Indeed, NeoGenomics is Natera's only competitor in this market. *See supra* note 3, at 4. Analysts believe NeoGenomics' RaDaR will see significant growth in the industry. Doc. 92-1 at 5, 20. For many of

NeoGenomics' sales, Natera will lose out on potential customers, profits, business relationships, and clinical opportunities. *See Douglas Dynamics*, 717 F.3d at 1345 (holding infringer's increase in market share more relevant than patentee's ability to maintain market share).

In this industry, biopharmaceutical partnerships are important. Doc. 7 at ¶¶ 103–04. By participating in clinical trials, companies can generate product data, gain credibility in the marketplace, and contribute to published research. *Id.* at ¶¶ 79, 103. Data from clinical trials also creates the support needed for insurance coverage determinations and entry into the larger clinical marketplace. *Id.* at ¶¶ 77, 103. If forced to compete against RaDaR for participation in future clinical studies, Natera could lose out on partnerships that substantially impact Signatera's future success, a loss that is challenging to quantify. *Id.* at ¶ 105; *see also Metalcraft of Mayville, Inc. v. The Toro Co.*, 848 F.3d 1358, 1368 (Fed. Cir. 2017) (holding that "[w]here the injury cannot be quantified" and "no amount of money damages is calculable," the harm is irreparable).

Natera has a first mover advantage as a pioneer in the tumor informed MRD market. Doc. 7 at ¶ 134. This advantage includes a period of exclusivity, in which Natera can establish brand recognition, customer loyalty, and business foundations. *Id.* at ¶ 136. Natera has never licensed the '035 patent, Doc. 14 at ¶ 16, and it has a right to exclusivity as the patent holder. *See Douglas Dynamics*, 717 F.3d at 1345; *Presidio*, 702 F.3d at 1363 (finding unwillingness to license supported irreparable injury). Natera's position as first mover will be unfairly cut short if RaDaR remains on the market.

NeoGenomics contends that the larger MRD market is the relevant market, and that Guardant Health is a bigger competitor for Natera than NeoGenomics. Doc. 89 at 23–24. While Guardant Health does operate in the MRD testing space, Doc. 92-1 at 26, it offers a tumor naïve product. *See* Doc. 7 at ¶ 53. Market analysts have confirmed that tumor informed tests are preferred by oncologists, Doc. 92-1 at 4, 17, making it highly likely that for many oncologists, the only two products are Natera's test and NeoGenomics' likely infringing test.

Even viewed as part of the larger MRD market, there is still irreparable harm. As the Federal Circuit has noted, "the existence of a two-player market" supports the granting of an injunction "because it creates an inference that an infringing sale amounts to a lost sale for the patentee," but "the converse is not automatically true." *Robert Bosch LLC v. Pylon Mfg. Corp.*, 659 F.3d 1142, 1151 (Fed. Cir. 2011) (cleaned up).

NeoGenomics contends that Natera has not shown that it has lost any contracts to NeoGenomics. But Natera has shown Moderna used RaDaR in at least one clinical study, Doc. 14-7 at 2, 11, and that NeoGenomics' representatives are promoting RaDaR to Natera's customers. Doc. 18-6. There is not abundant evidence of lost sales, but RaDaR is relatively new to the market; it has only been commercially available since March, Doc. 14-4, and it only received approval for Medicare coverage for use with certain specified cancers in July. Doc. 1-30. Competition and potential lost sales from RaDaR are likely to occur and cause harm.

Finally, NeoGenomics contends that Natera unreasonably delayed in filing this action. *See* Doc. 89 at 20–22. Natera received this patent in December 2022, Doc. 1-2 at

2, NeoGenomics entered the clinical market in March 2023, Doc. 14-4, and this lawsuit was filed in July 2023. Doc. 1. In the interim, Natera was already involved in ongoing patent infringement litigation over related patents. *See ArcherDX, Inc.*, No. 20-CV-125.

This is not undue delay. A patentee is not required to "sue all infringers at once" and suing four months after an infringer enters the market is relatively quick. *Robert Bosch*, 659 F.3d at 1151. Moreover, "picking off one infringer at a time is not inconsistent with being irreparably harmed." *Id.* (citing *Pfizer, Inc. v. Teva Pharms. USA, Inc.*, 429 F.3d 1364, 1381 (Fed. Cir. 2005)).

Natera has shown a likelihood of irreparable harm.

#### **B.** Causal Nexus

In addition to likely irreparable harm, the patentee must also demonstrate "that a sufficiently strong causal nexus relates the alleged harm to the alleged infringement." *Bio-Rad Lab'ys, Inc. v. 10X Genomics, Inc.*, 967 F.3d 1353, 1377–78 (Fed. Cir. 2020) (cleaned up); *Apple Inc. v. Samsung Elecs. Co., Ltd.*, 809 F.3d 633, 640 (Fed. Cir. 2015). While the infringing feature does not have to be the sole reason a product is bought, *Apple*, 809 F.3d at 641–42, a causal nexus exists if "the infringing feature drives consumer demand for the accused product." *TEK Glob.*, 920 F.3d at 792 (quoting *Apple Inc. v. Samsung Elecs. Co.*, 695 F.3d 1370, 1375–76 (Fed. Cir. 2012)).

There is a causal nexus between the likely infringement and harm. It appears highly likely that NeoGenomics' predecessor built RaDaR using the methods of the '035 patent as a foundation. As previously discussed *supra*, the likely infringement allows NeoGenomics to offer RaDaR as a tumor informed MRD assay, doctors who order these

tests often prefer tumor informed tests, and RaDaR's ability to perform tumor informed testing is what drives consumer demand for it.

# C. BALANCE OF EQUITIES

The third factor requires the patentee show that "the balance of equities tips in its favor." *BlephEx*, 24 F.4th at 1398. When balancing equities, the court "weigh[s] the harm to the moving party if the injunction is not granted against the harm to the non-moving party if the injunction is granted." *Metalcraft*, 848 F.3d at 1369 (citing *Hybritech Inc. v. Abbott Lab'ys*, 849 F.2d 1446, 1457 (Fed. Cir. 1988)). The court "may consider the parties' sizes, products, and revenue sources" in its analysis, but the "expenses incurred in creating the infringing products and the consequences of its infringement are irrelevant." *Bio-Rad Lab'ys*, 967 F.3d at 1378 (cleaned up).

The balance of equities weighs in favor of granting a preliminary injunction. A 2020 analyst report identified Signatera as Natera's "most valuable offering." Doc. 7 at ¶ 158. In 2023, that same firm predicted that advances related to Signatera were key to Natera's future success. *See* Doc. 7-18 at 4 (section titled "Key catalysts for Natera's stock from here").

A May 2023 report predicts revenue from Signatera will increase from \$130.4 million in 2022 to \$432.4 million in 2025 and make up 52.1% of Natera's total growth in revenue. Doc. 7 at ¶¶ 156–57; Doc. 7-18 at 6. In the first quarter of 2023, the "vast majority" of Natera's revenue attributable to oncology products came from "clinical Signatera volume growth." Doc. 7-18 at 4 (section titled "More on Signatera in greater detail"). Signatera is important to Natera's economic success.

In contrast, NeoGenomics is not as dependent on RaDaR's success. It has "the broadest cancer diagnostic testing menu in the U.S.," with over 600 tests related to cancer diagnostics; NeoGenomics launches roughly 60 to 70 new test products each year. Doc. 7-19 at 3 (April 2023 report, section titled "Investment Thesis").

RaDaR also only recently became commercially available. *See* Doc. 14-4. Although NeoGenomics spent a significant amount of money acquiring Inivata and the RaDaR test, this expense does not tip the balance of equities in favor of NeoGenomics. *See Bio-Rad Lab'ys*, 967 F.3d at 1378.

The harm to Natera if a preliminary injunction is not granted outweighs the harm to NeoGenomics in granting the injunction. *See Metalcraft*, 848 F.3d at 1369. Signatera has been on the market longer and is predicted to be a major contributor to Natera's future success. In comparison, RaDaR is relatively new to the market and is not a major product in NeoGenomics' portfolio. Natera's projected revenue streams and dependence on Signatera tips the balance of equities in favor of granting the preliminary injunction. *See Bio-Rad Lab'ys*, 967 F.3d at 1378.

#### D. PUBLIC INTEREST

It is in the public's interest to uphold patent rights. See i4i Ltd. P'ship v. Microsoft Corp., 598 F.3d 831, 863 (Fed. Cir. 2010); Douglas Dynamics, 717 F.3d at 1346 (stating the public has a "general interest in the judicial protection of property rights in inventive technology"). Before granting an injunction, courts must balance "protecting the patentee's rights" with any adverse effects on the public. i4i, 598 F.3d at 863; see also Metalcraft, 848 F.3d at 1369 ("[T]he district court should focus on whether a critical

public interest would be injured by the grant of injunctive relief."). To reduce any adverse effects on the public interest, an injunction can be crafted to exclude current users of the enjoined product. *See i4i*, 598 F.3d at 863.

The public interest in enforcing patent rights tips in favor of granting a preliminary injunction. Anyone in need of a tumor informed MRD test will be able to get one from Natera; Signatera is clinically validated for use with the same cancers as RaDaR. *See* Doc. 10-39 at 2–3 (RaDaR cancer coverage); Doc. 10-5 at 2 (Signatera's Medicare coverage); Doc. 10-15 at 13 (listing Signatera's published indications as of 2022). Natera has the capacity to take on more customers. *See, e.g.*, Doc. 10-13 at 7 (Form 10-K for 2021 discussing Natera's "global network of over 100 laboratory and distribution partners"); Doc. 10-17 at 17 (2022 presentation discussing Natera's ability to scale up); Doc. 14 at ¶ 14; *see also* Doc. 7 at ¶¶ 89–92. The injunction can be crafted in a way that does not disrupt clinical trials and ongoing research and so that current patients can continue to use RaDaR. *See* discussion *infra*.

Consumer choice is important, as NeoGenomics points out. Doc. 89 at 30–31. But competition from an infringing product does not benefit the public, and it impedes innovation. *See Douglas Dynamics*, 717 F.3d at 1346 (so holding after verdict in favor of patent holder). While this is not a final decision on the merits, Natera has made a strong case for infringement and validity of the '035 patent. The need to protect consumer choice does not weigh heavily in favor of denying an injunction, in light of other evidence. *See Shure, Inc. v. ClearOne, Inc.*, No. 17-CV-3078, 2019 WL 3555098, at \*25 (N.D. Ill. Aug. 5, 2019).

As NeoGenomics points out, there are patients currently using RaDaR who would be harmed if it were withdrawn from the market, and there are clinical trials and research projects, which depend on the use of RaDaR, in process or approved to begin. *See* Doc.

89 at 31. The public interest does not support enjoining these uses, despite the potential

for infringement.

The public generally benefits from clinical trials, which should be carried out if they have been approved and completed if they are in process. The same is true for research projects already in process. And, as Natera recognizes, current patients using RaDaR, whether in clinical trials or otherwise, cannot at this point in their medical care use Signatera as a substitute. *See* Doc. 5 at 31 (asking for an injunction that allows patients currently using RaDaR to continue use); Doc. 139 at 19. The Court agrees that

avoiding disruption to ongoing treatment, research, and clinical studies is proper, and the

E. CONCLUSION

preliminary injunction will be crafted to protect those interests.

Natera has shown that it is likely to succeed on the merits, it is likely to suffer irreparable harm in the absence of preliminary relief, the balance of equities tips in its favor, and an injunction is in the public interest. A preliminary injunction is appropriate.

It is **ORDERED** that Natera's preliminary injunction motion, Doc. 5, is **GRANTED.** The preliminary injunction will issue separately.

This the 27th day of December, 2023.

UNITED STATES DISTRICT JUDGE

# IN THE UNITED STATES DISTRICT COURT FOR THE MIDDLE DISTRICT OF NORTH CAROLINA

NATERA, INC.,	)
Plaintiff,	)
V.	) ) 1:23-CV-629
	)
NEOGENOMICS LABORATORIES, INC.,	)
Defendant.	)

# **PRELIMINARY INJUNCTION**

Plaintiff Natera, Inc. moves for a preliminary injunction based on alleged infringement of two of its patents by defendant NeoGenomics Laboratories Inc. On consideration of the entire record, and as set forth in more detail in an order entered concurrently, the Court finds and concludes that Natera has demonstrated that it is likely to succeed on the merits of its claim. It is likely (1) NeoGenomics is infringing U.S. Patent No. 11,519,035 by making and selling the RaDaR assay, and NeoGenomics has not presented a substantial question of the validity of the asserted claims of the '035 patent; (2) Natera is substantially likely to suffer irreparable harm from NeoGenomics' ongoing infringement of the '035 patent; and (3) the balance of equities and the public interest favor protecting Natera and granting a preliminary injunction.

#### IT IS HEREBY ORDERED:

- 1. Natera's Motion for a Preliminary Injunction, Doc. 5, is **GRANTED**.
- 2. NeoGenomics and its privies, subsidiaries, affiliates, officers, agents, servants, employees, attorneys, and those acting or attempting to act in concert or

participation with them, are **ENJOINED** from (1) making, using, selling, or offering for sale in the United States the (a) the accused RaDaR assay ("the Accused Assay") or (b) any assay or product not more than colorably different from the Accused Assay whose manufacture, use, importation, offer for sale, or sale would infringe any asserted claim of U.S. Patent No. 11,519,035; and (2) promoting, advertising, marketing, servicing, distributing, or supplying the above Accused Assay so as to induce others' infringement.

- 3. Notwithstanding the foregoing, NeoGenomics may continue to make, use, and sell the Accused Assay solely for continued use of the Accused Assay
  - a. for those patients already using it before the entry of this injunction,
  - b. in support of research and development with other persons or entities on projects or studies that began before the entry of this injunction, or
  - c. for use in or in support of clinical trials in process or already approved by an agency of the United States.
- 4. NeoGenomics **SHALL**, within 10 days from the date of issuance of this Preliminary Injunction, provide notice and a copy of this Preliminary Injunction to (1) each of its privies, subsidiaries, affiliates, officers, agents, servants, employees, and attorneys; and (2) any other person or entity acting or attempting to act in concert or participation with NeoGenomics with respect to any of the enjoined activities, such that above persons and entities are duly noticed and bound by this Order under Federal Rule of Civil Procedure 65(d)(2). NeoGenomics shall further provide proof of each such notice to this

# 42xqqA

Court by filing proof of service within 14 days from the date of issuance of this Preliminary Injunction.

This the 27th day of December, 2023.

UNITED STATES DISTRICT JUDGE

# **EXHIBIT 2**



# (12) United States Patent Rabinowitz et al.

# (10) Patent No.: US 11,519,035 B2 (45) Date of Patent: \*Dec. 6, 2022

#### (54) METHODS FOR SIMULTANEOUS AMPLIFICATION OF TARGET LOCI

(71) Applicant: Natera, Inc., San Carlos, CA (US)

(72) Inventors: Matthew Rabinowitz, San Francisco, CA (US); Matthew Micah Hill, Belmont, CA (US); Bernhard Zimmermann, Manteca, CA (US); Johan Baner, San Francisco, CA (US); George Gemelos, Portland, OR (US); Milena Banjevic, Los Altos Hills, CA (US); Allison Ryan, Belmont, CA (US); Styrmir Sigurjonsson, San Jose, CA (US); Zachary Demko, San Francisco,

CA (US)

(73) Assignee: Natera, Inc., San Carlos, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35

U.S.C. 154(b) by 0 days.

This patent is subject to a terminal dis-

claimer.

(21) Appl. No.: 16/934,407

(22) Filed: Jul. 21, 2020

(65) **Prior Publication Data** 

US 2020/0347454 A1 Nov. 5, 2020

#### Related U.S. Application Data

(63) Continuation of application No. 16/743,724, filed on Jan. 15, 2020, now Pat. No. 10,731,220, which is a continuation of application No. 16/399,268, filed on Apr. 30, 2019, now Pat. No. 10,538,814, which is a continuation of application No. 16/140,298, filed on Sep. 24, 2018, which is a continuation of application No. 14/918,544, filed on Oct. 20, 2015, now Pat. No. 10,316,362, which is a continuation-in-part of application No. 14/877,925, filed on Oct. 7, 2015, now abandoned, and a continuation-in-part of application No. 14/692,703, filed on Apr. 21, 2015, now Pat. No. 10,179,937, and a continuation-in-part of application No. 14/538,982, filed on Nov. 24, 2014, now Pat. No. 9,677,118, said application No. 14/877,925 is a continuation-in-part of application No. 14/225,356, filed on Mar. 25, 2014, now abandoned, and a continuation-in-part of application No. 13/780,022, filed on Feb. 28, 2013, now abandoned, and a continuation of application No. 13/683,604, filed on Nov. 21, 2012, now abandoned, said application No. 14/225,356 is a continuation of application No. PCT/US2012/058578, filed on Oct. 3, 2012, said application No. 13/780,022 is a continuation-in-part of application No. 13/683,604, filed on Nov. 21, 2012, now abandoned, and a continuation-in-part application of PCT/US2012/058578, filed on Oct. 3, 2012, and a continuation-in-part of application No. 13/335,043, filed on Dec. 22, 2011, now Pat. No. 10,113,196, and

a continuation-in-part of application No. 13/300,235, filed on Nov. 18, 2011, now Pat. No. 10,017,812, and a continuation-in-part of application No. 13/110,685, filed on May 18, 2011, now Pat. No. 8,825,412, said (Continued)

(51)	Int. Cl.	
	C12Q 1/68	(2018.01)
	C12Q 1/6883	(2018.01)
	C12Q 1/6811	(2018.01)
	C12Q 1/6848	(2018.01)
	C12Q 1/6809	(2018.01)
	C12Q 1/6851	(2018.01)
	C12Q 1/6844	(2018.01)
	C12Q 1/6869	(2018.01)
	C12Q 1/6855	(2018.01)
	C12O 1/6874	(2018.01)

(52) U.S. Cl.

(58) Field of Classification Search

None

See application file for complete search history.

#### (56) References Cited

# U.S. PATENT DOCUMENTS

4,040,785 A 8/1977 Kim et al. 4,683,195 A 7/1987 Mullis et al. (Continued)

## FOREIGN PATENT DOCUMENTS

CN 1650032 A 8/2005 CN 1674028 A 9/2005 (Continued)

#### OTHER PUBLICATIONS

US 8,501,409 B2, 08/2013, Simen et al. (withdrawn) (Continued)

Primary Examiner — Stephanie K Mummert

#### (57) ABSTRACT

The invention provides methods for simultaneously amplifying multiple nucleic acid regions of interest in one reaction volume as well as methods for selecting a library of primers for use in such amplification methods. The invention also provides library of primers with desirable characteristics, such as minimal formation of amplified primer dimers or other non-target amplicons.

15 Claims, 58 Drawing Sheets

Specification includes a Sequence Listing.

## US 11,519,035 B2

Page 2

6,214,558 B1

6,221,603 B1 6,258,540 B1

6,300,077 B1

6,329,179 B1 6,335,167 B1 6,440,706 B1

6,479,235 B1

4/2001

4/2001 7/2001

10/2001

12/2001

1/2002

8/2002

11/2002

Shuber et al.

Shuber et al.

Kopreski Pinkel et al.

Vogelstein et al.

Schumm et al.

Mahtani

Lo et al

#### Related U.S. Application Data

application No. 13/683,604 is a continuation-in-part of application No. 13/300,235, filed on Nov. 18, 2011, now Pat. No. 10,017,812, and a continuation-in-part of application No. 13/110,685, filed on May 18, 2011, now Pat. No. 8,825,412, said application No. PCT/US2012/058578 is a continuation-in-part of application No. 13/300,235, filed on Nov. 18, 2011, now Pat. No. 10,017,812, said application No. 13/335,043 is a continuation-in-part of application No. 13/300,235, filed on Nov. 18, 2011, now Pat. No. 10,017,812, and a continuation-in-part of application No. 13/110,685, filed on May 18, 2011, now Pat. No. 8,825,412, said application No. 13/310,235 is a continuation-in-part of application No. 13/110,685, filed on May 18, 2011, now Pat. No. 8,825,412, now Pat. No. 8,825,412.

(60) Provisional application No. 62/148,173, filed on Apr. 14, 2015, provisional application No. 62/146,188, filed on Apr. 10, 2015, provisional application No. 62/147,377, filed on Apr. 14, 2015, provisional application No. 62/066,514, filed on Oct. 21, 2014, provisional application No. 61/994,791, filed on May 16, 2014, provisional application No. 61/987,407, filed on May 1, 2014, provisional application No. 61/982,245, filed on Apr. 21, 2014, provisional application No. 61/634,431, filed on Feb. 29, 2012, provisional application No. 61/675,020, filed on Jul. 24, 2012, provisional application No. 61/683,331, filed on Aug. 15, 2012, provisional application No. 61/542,508, filed on Oct. 3, 2011, provisional application No. 61/426,208, filed on Dec. 22, 2010, provisional application No. 61/571,248, filed on Jun. 23, 2011, provisional application No. 61/516,996, filed on Apr. 12, 2011, provisional application No. 61/448,547, filed on Mar. 2, 2011, provisional application No. 61/462,972, filed on Feb. 9, 2011, provisional application No. 61/398,159, filed on Jun. 21, 2010, provisional application No. 61/395,850, filed on May 18, 2010.

### (56) References Cited

### U.S. PATENT DOCUMENTS

4,942,124 A 5,486,477 A 7/1990 Church et al. 1/1996 Carver 5,635,366 A 6/1997 Cooke et al. 5,648,220 A 5,716,776 A 7/1997 2/1998 Bianchi et al. Bogart 5,753,467 A 5/1998 Jensen et al. 5,824,467 A 10/1998 Mascarenhas 5.854,033 A 12/1998 Lizardi 5,860,917 A 1/1999 Comanor et al. 5,891,734 A 4/1999 Gill et al. 5,952,170 A 5,962,223 A 9/1999 Stroun et al. Whiteley et al. Hyland et al. 10/1999 5,972,602 11/1999 5,976,790 A 11/1999 Pinkel et al. 5,994,148 A 6,001,611 A Stewart et al. Will 11/1999 12/1999 6,025,128 A 2/2000 Veltri et al. 6,066,454 A 5/2000 Lipshutz et al. 6,100,029 A 8/2000 Lapidus et al. 6,108,635 A 8/2000 Herren et al. 6,124,120 A 9/2000 Lizardi 6,143,496 A 11/2000 Brown et al. 12/2000 6.156.504 A Gocke et al. 6,180,349 B1 1/2001 Ginzinger 6,235,472 B1 2/2001 Landegren et al. 6,489,135 B1 6,605,451 B1 12/2002 Parrott et al. 8/2003 Marmaro et al. 6,617,137 B2 9/2003 Dean et al. 4/2004 5/2004 6,720,140 B1 Hartley et al 6.858.412 B2 Willis et al. Goldsborough 6,794,140 B1 9/2004 6,807,491 B2 10/2004 Pavlovic et al. 6,927,028 B2 6,852,487 B1 8/2005 10/2005 Lo et al. Barany et al. 6,958,211 B2 10/2005 Vingerhoets et al. 6,964,847 B1 7,035,739 B2 11/2005Englert 4/2006 Schadt et al 7,058,517 B1 6/2006 Denton et al 7,058,616 B1 6/2006 Larder et al. 7,101,663 B2 7,153,656 B2 9/2006 12/2006 Godfrey et al Nolan et al. 7,218,764 B2 5/2007 Vaisberg et al 7,297,485 B2 7,332,277 B2 7,410,764 B2 11/2007 Bornarth et al. 2/2008 Dhallan 8/2008 Gocke et al 7,414,118 B1 8/2008 Mullah et al. 7,442,506 B2 7,459,273 B2 12/2008 12/2008 Dhallan Jones et al 7,645,576 B2 1/2010 Lo et al. 7,655,399 B2 7,700,325 B2 2/2010 Cantor et al 4/2010 Cantor et al. 7,718,367 B2 5/2010 Lo et al. 7,718,370 B2 6/2010 Dhallan 7,727,720 B2 7,741,463 B2 6/2010Dhallan 6/2010 9/2010 Gormley et al. 7,790,418 B2 Mayer 7,805,282 B2 9/2010 Casey et al 7,838,647 B2 7,981,609 B2 11/2010 7/2011 Hahn et al. Rubin et al 7,888,017 B2 8/2011 Quake 8,008,018 B2 9/2011Quake et al 8,024,128 B2 8,133,719 B2 9/2011 Rabinowitz 3/2012 Drmanac et al. 8,137,912 B2 3/2012 Kapur et al. 8,173,370 B2 8,168,389 B2 8,236,503 B2 5/2012 6/2012 Oeth et al. Shoemaker et al. 8/2012 Faham et al. 8,195,415 B2 10/2012 Fan et al 8,296,076 B2 8,304,187 B2 11/2012 11/2012 Fan et al. Fernando 8,318,430 B2\* 11/2012 Chuu ...... C12Q 1/6883 8,318,434 B2 8,323,897 B2 8,372,584 B2 11/2012 12/2012 Cuppens et al. Andersen et al 2/2013 Shoemaker et al. 8,389,557 B2 8,389,578 B2 3/2013 Singh et al. 3/2013 5/2013 Went et al. Dube et al. 8,450,063 B2 8,467,976 B2 8,515,679 B2 8,532,930 B2 8/2013 Lo et al. 8/2013 Rabinowitz et al. 9/2013 Rabinowitz et al. 8,532,930 B2 8,679,741 B2 8,682,592 B2 8,703,652 B2 8,706,422 B2 3/2014 Hoyal-Wrightson et al. 3/2014 Rabinowitz et al.  $\frac{4/2014}{4/2014}$ Quake et al. Lo et al. 8,748,103 B2 Faham et al 6/2014 8,822,153 B2 9/2014 Hayes et al. 8,825,412 B2 9,005,894 B2 9,051,602 B2 9/2014 Rabinowitz et al. 4/2015 Ladner et al. 6/2015 Oliphant et al. 9,085,798 B2 7/2015 Chee 9,206,475 B2 12/2015 Gerdes et al. 9,228,234 B2 1/2016 Rabinowitz et al. 9,323,888 B2 4/2016 Rava et al 9,404,150 B2 8/2016 Lee et al

8/2016 Rabinowitz et al.

9,424,392 B2

# US 11,519,035 B2

(56)	Referen	nces Cited	2005/0009069 A1 2005/0043894 A1		Liu et al. Fernandez
U	.S. PATENT	DOCUMENTS	2005/0049793 A1	3/2005	Paterlini-brechot
0.452.257 B	0/2016	Havel Wrightson et al	2005/0053950 A1 2005/0064476 A1		Ubani et al. Huang et al.
9,453,257 E 9,476,095 E		Hoyal-Wrightson et al. Vogelstein et al.	2005/0079521 A1	4/2005	Beaulieu et al.
9,487,829 E	11/2016	Vogelstein et al.	2005/0079535 A1		Kirchgesser et al. Katz et al.
9,493,828 E 9,506,119 E		Rava et al. Faham et al.	2005/0123914 A1 2005/0130173 A1		Leamon et al.
9,598,731 E		Talasaz	2005/0142577 A1	6/2005	Jones et al.
9,677,118 E		Zimmermann et al.	2005/0144664 A1 2005/0164241 A1	6/2005	Smith et al. Hahn et al.
10,017,810 E 10,041,127 E		Iafrate et al. Talasaz	2005/0164241 A1 2005/0164252 A1	7/2005	
10,061,890 E		Rabinowitz et al.	2005/0216207 A1		Kermani
10,081,839 E		Rabinowitz et al.	2005/0221341 A1 2005/0227263 A1		Shimkets et al. Green et al.
10,083,273 E 10,174,369 E		Rabinowitz et al. Rabinowitz et al.	2005/0250111 A1		Xie et al.
10,179,937 E	1/2019	Babiarz et al.	2005/0255508 A1		Casey et al.
10,227,652 E		Rabinowitz et al. Rabinowitz et al.	2005/0272073 A1 2005/0282185 A1		Vaisberg et al. Lo et al.
10,240,202 E 10,260,096 E		Rabinowitz et al.	2006/0019278 A1	1/2006	Lo et al.
10,266,893 E	4/2019	Rabinowitz et al.	2006/0040300 A1 2006/0046258 A1		Dapprich et al. Lapidus et al.
10,308,981 E 10,316,362 E		Sparks et al. Babiarz et al.	2006/0040238 A1 2006/0051799 A1		Iwaki et al.
10,351,906 B		Zimmermann et al.	2006/0052945 A1		Rabinowitz et al.
10,392,664 E		Rabinowitz et al.	2006/0057618 A1 2006/0068394 A1		Piper et al. Langmore et al.
10,450,597 E 10,522,242 E		Iafrate et al. Rabinowitz et al.	2006/0008354 A1 2006/0088574 A1		Manning et al.
10,526,658 E		Babiarz et al.	2006/0094010 A1	5/2006	Giles et al.
10,538,814 B		Babiarz et al.	2006/0099614 A1 2006/0121452 A1		Gill et al. Dhallan
10,557,172 E 10,597,708 E		Babiarz et al. Zimmermann et al.	2006/0121432 A1 2006/0134662 A1		Pratt et al.
10,597,709 E	3/2020	Zimmermann et al.	2006/0141499 A1		Sher et al.
10,597,723 E		Babiarz et al.	2006/0229823 A1 2006/0210997 A1	8/2006 9/2006	Myerson et al.
10,655,180 E 10,711,309 E		Babiarz et al. Rabinowitz et al.	2006/0216153 A1	9/2006	Wobben et al.
10,731,220 E	8/2020	Babiarz et al.	2006/0216738 A1		Wada et al.
10,774,380 E 10,793,912 E		Ryan et al. Babiarz et al.	2006/0228721 A1 2006/0248031 A1		Leamon et al. Kates et al.
10,793,912 E 10,894,976 E		Stray et al.	2006/0281105 A1	12/2006	Li et al.
11,111,543 E	9/2021	Rabinowitz et al.	2006/0292599 A1		Ritz et al.
11,111,544 E 11,111,545 E		Rabinowitz et al. Babiarz et al.	2007/0020640 A1 2007/0027636 A1		McCloskey et al. Rabinowitz
11,111,545 E		Quake et al.	2007/0031857 A1	2/2007	Makarov et al.
2001/0051341 A		Lo et al.	2007/0037166 A1 2007/0042384 A1		Wohlgemuth et al. Li et al.
2001/0053519 A 2002/0006622 A		Fodor et al. Bradley et al.	2007/0059700 A1		Tao et al.
2002/0107640 A		Ideker et al.	2007/0059707 A1		Cantor et al.
2002/0119478 A		Umansky et al.	2007/0122805 A1 2007/0128624 A1		Cantor et al. Gormley et al.
2002/0182622 A 2003/0009295 A		Nakamura et al. Markowitz et al.	2007/0134658 A1	6/2007	Bohmer et al.
2003/0040620 A	1 2/2003	Langmore et al.	2007/0178478 A1		Dhallan
2003/0044388 A 2003/0065535 A		Lo et al. Karlov et al.	2007/0178501 A1 2007/0184467 A1		Rabinowitz et al. Rabinowitz et al.
2003/0003333 A		Pavlovic et al.	2007/0202525 A1	8/2007	Quake et al.
2003/0087276 A	1 5/2003	Kopreski	2007/0202536 A1 2007/0207466 A1		Yamanishi et al. Cantor et al.
2003/0101000 A 2003/0119004 A		Bader et al. Wenz et al.	2007/0207400 A1 2007/0212689 A1		Bianchi et al.
2003/0119004 A		Gill et al.	2007/0243549 A1	10/2007	Bischoff
2003/0211489 A		Shen et al.	2007/0259351 A1 2008/0020390 A1	1/2007	Chinitz Mitchell
2003/0228613 A 2003/0232348 A		Bornarth et al. Jones et al.	2008/0026390 A1	1/2008	
2003/0232353 A	1 12/2003	Kennedy et al.	2008/0038733 A1		Bischoff et al.
2004/0009518 A		Lo et al. Threadgill et al.	2008/0050739 A1 2008/0070792 A1	2/2008 3/2008	Stoughton et al. Stoughton
2004/0033596 A 2004/0067493 A		Matsuzaki et al.	2008/0071076 A1		Hahn et al.
2004/0096874 A	1 5/2004	Neville et al.	2008/0085836 A1	4/2008	
2004/0115629 A 2004/0117346 A		Panzer et al. Stoffel et al.	2008/0090239 A1 2008/0096766 A1	4/2008 4/2008	
2004/0117346 A 2004/0126760 A		Broude	2008/0102455 A1	5/2008	Poetter
2004/0137470 A	1 7/2004	Dhallan et al.	2008/0138809 A1		Kapur et al.
2004/0146866 A 2004/0157243 A		Fu Huang et al.	2008/0182244 A1 2008/0193927 A1	7/2008 8/2008	
2004/0137243 A 2004/0185495 A		Schueler et al.	2008/0193927 A1 2008/0220422 A1	9/2008	
2004/0197797 A	1 10/2004	Inoko et al.	2008/0234142 A1	9/2008	Lietz
2004/0209299 A		Pinter et al.	2008/0243398 A1		Rabinowitz et al.
2004/0229231 A 2004/0236518 A		Frudakis et al. Pavlovic et al.	2008/0280292 A1 2008/0286783 A1		Wangh et al. Hosono et al.
2004/0259100 A		Gunderson et al.	2008/0299562 A1		Oeth et al.

# US 11,519,035 B2

(56)	Referen	ces Cited		2012/0010085 A 2012/0028814 A		Rava et al. Toloue et al.
U.S.	PATENT	DOCUMENTS		2012/0028814 A 2012/0034603 A		
				2012/0034685 A		Sparks et al.
2008/0305473 A1		Chowdary et al.		2012/0108460 A 2012/0122701 A		Quake et al. Ryan et al.
2009/0023190 A1 2009/0029377 A1		Lao et al. Lo et al.		2012/0122701 A 2012/0165203 A		Quake et al.
2009/0053719 A1		Lo et al.		2012/0185176 A		Rabinowitz et al.
2009/0087847 A1		Lo et al.		2012/0190020 A 2012/0190021 A		Oliphant et al. Oliphant et al.
2009/0098534 A1 2009/0099041 A1		Weier et al. Church et al.		2012/0190521 A		Oliphant et al.
2009/0143570 A1		Jiang et al.		2012/0191358 A		Oliphant et al.
2009/0176662 A1		Rigatti et al.		2012/0196754 A 2012/0208706 A		Quake et al. Downing et al.
2009/0221620 A1 2009/0228299 A1		Luke et al. Kangarloo et al.		2012/0208/00 A 2012/0214678 A		Rava et al.
2009/0263800 A1		Wohlgemuth et al.		2012/0264121 A		Rava et al.
2009/0280479 A1		Hoon et al.		2012/0264618 A 2012/0270212 A		Nygren Rabinowitz et al.
2009/0317817 A1 2010/0035232 A1		Oeth et al. Ecker et al.		2012/0270212 A		Rava et al.
2010/0041048 A1		Diehl et al.		2012/0295810 A		Quake et al.
2010/0086914 A1		Bentley et al.		2012/0295819 A 2013/0017549 A		Leamon et al.
2010/0105049 A1 2010/0112575 A1*		Ehrich et al.	C12O 1/6883	2013/0017349 A 2013/0022973 A		Hansen et al.
2010/0112373 A1	3/2010	Tan	435/6.11	2013/0024127 A	.1 1/2013	Stuelpnagel
2010/0112586 A1		Stoughton et al.		2013/0034546 A 2013/0040375 A		Rava et al. Sparks et al.
2010/0112590 A1		Lo et al.		2013/0040373 A 2013/0060483 A		
2010/0120038 A1 2010/0124751 A1		Mir et al. Quake et al.		2013/0069869 A	.1 3/2013	Akao et al.
2010/0129792 A1	5/2010	Makrigiorgos		2013/0085681 A		Deciu et al. Sparks et al.
2010/0129874 A1		Mitra et al. Fan et al.		2013/0090250 A 2013/0116130 A		Fu et al.
2010/0138165 A1 2010/0171954 A1		Quake et al.		2013/0123120 A	.1 5/2013	Zimmermann et al.
2010/0173394 A1	7/2010	Colston et al.		2013/0130923 A		Ehrich et al. Pushkarev et al.
2010/0184043 A1		Mitchell et al.		2013/0157870 A 2013/0172211 A		Oliphant et al.
2010/0184069 A1 2010/0184152 A1		Fernando et al. Sandler		2013/0178373 A	.1 7/2013	Rabinowitz et al.
2010/0196892 A1	8/2010	Quake et al.		2013/0190653 A		Alvarez Ramos
2010/0203538 A1		Dube et al.		2013/0196862 A 2013/0210644 A		Rabinowitz et al. Stoughton et al.
2010/0216151 A1 2010/0216153 A1		Lapdus et al. Lapidus et al.		2013/0225422 A	.1 8/2013	Rabinowitz et al.
2010/0248231 A1	9/2010	Wei et al.		2013/0237431 A		Lo et al. Rabinowitz et al.
2010/0255492 A1		Quake et al.		2013/0252824 A 2013/0253369 A		Rabinowitz et al.
2010/0256013 A1 2010/0273159 A1	10/2010	Quake et al. Melo		2013/0261004 A	.1 10/2013	Ryan et al.
2010/0273219 A1	10/2010	May et al.		2013/0274116 A		Rabinowitz et al.
2010/0273678 A1		Alexandre et al.		2013/0288252 A 2013/0303461 A		Sparks et al. Iafrate et al.
2010/0285478 A1 2010/0285537 A1		Chen et al. Zimmermann		2013/0310260 A	.1 11/2013	Kim et al.
2010/0291572 A1	11/2010	Stoughton et al.		2013/0323731 A		Lo et al.
2010/0291635 A1	11/2010	Peleg Lo et al.		2013/0325360 A 2014/0032128 A		Deciu et al. Rabinowitz et al.
2010/0323352 A1 2011/0015096 A1	1/2010			2014/0038830 A	.1 2/2014	Srinivasan et al.
2011/0033862 A1		Rabinowitz et al.		2014/0051585 A 2014/0065621 A		Prosen et al. Mhatre et al.
2011/0039724 A1 2011/0045462 A1		Lo et al. Fu et al.		2014/0065021 A 2014/0066317 A		Talasaz
2011/0043402 A1 2011/0064824 A1		Lascoste et al.		2014/0087385 A		Rabinowitz et al.
2011/0071031 A1		Khripin et al.		2014/0094373 A 2014/0100126 A		Zimmermann et al. Rabinowitz
2011/0086769 A1 2011/0092763 A1		Oliphant et al. Rabinowitz et al.		2014/0100120 A 2014/0100134 A		Rabinowitz et al.
2011/0092703 A1 2011/0105353 A1		Lo et al.		2014/0106975 A		Stoughton et al.
2011/0130558 A1		Ritt et al.		2014/0141981 A 2014/0154682 A		Zimmermann et al. Rabinowitz et al.
2011/0151442 A1 2011/0159499 A1		Fan et al. Hindson et al.		2014/0155274 A		Xie et al.
2011/0139499 A1 2011/0160078 A1		Fodor et al.		2014/0162269 A		Rabinowitz
2011/0178719 A1		Rabinowitz et al.		2014/0186827 A 2014/0193816 A		Pieprzyk et al. Rabinowitz et al.
2011/0201507 A1 2011/0212446 A1		Rava et al. Wang et al.		2014/0206552 A		Rabinowitz et al.
2011/0212440 A1 2011/0212846 A1	9/2011			2014/0227691 A		May et al.
2011/0224087 A1	9/2011	Quake et al.		2014/0227705 A 2014/0242588 A		Vogelstein et al. Van Den Boom et al.
2011/0246083 A1 2011/0251149 A1		Fan et al. Perrine et al.		2014/0242588 A 2014/0256558 A		Varley et al.
2011/0231149 A1 2011/0288780 A1		Rabinowitz et al.		2014/0256569 A	.1 9/2014	Rabinowitz et al.
2011/0294699 A1	12/2011	Lee et al.		2014/0272956 A		Huang et al.
2011/0300608 A1 2011/0301854 A1		Ryan et al. Curry et al.		2014/0274740 A 2014/0287934 A		Srinivasan et al. Szelinger et al.
2011/0301834 A1 2011/0312503 A1		Chuu et al.		2014/028/934 A 2014/0296081 A		Diehn et al.
2011/0318734 A1	12/2011	Lo et al.		2014/0329245 A	.1 11/2014	Spier et al.
2012/0003635 A1		Lo et al.		2014/0336060 A		Rabinowitz
2012/0003637 A1	1/2012	Lo et al.		2015/0051087 A	.1 2/2015	Rabinowitz et al.

# US 11,519,035 B2

(56)	Referen	nces Cited	2019/0211393			Rabinowitz et al. Rabinowitz et al.
U.S.	PATENT	DOCUMENTS	2019/0211399 2019/0211402		7/2019	
			2019/0211406			Babiarz et al.
2015/0064695 A1 2015/0086477 A1		Katz et al. Mitchell et al.	2019/0249241 2019/0256894			Rabinowitz et al. Zimmermann et al.
2015/0080477 A1 2015/0087535 A1	3/2015		2019/0256906		8/2019	Rabinowitz et al.
2015/0147815 A1		Babiarz et al.	2019/0256907 2019/0256908			Ryan et al. Rabinowitz et al.
2015/0167069 A1 2015/0197786 A1		Schutz et al. Osborne et al.	2019/0256909			Rabinowitz et al.
2015/0211050 A1		Iafrate et al.	2019/0256912			Rabinowitz et al.
2015/0218631 A1		Chuu et al.	2019/0256916 2019/0256917			Babiarz et al. Babiarz et al.
2015/0232938 A1 2015/0265995 A1		Mhatre Head et al.	2019/0256919			Babiarz et al.
2015/0299812 A1		Talasaz	2019/0256924			Vogelstein et al.
2015/0315657 A1		Rhodes et al.	2019/0256931 2019/0264277			Babiarz et al. Rabinowitz et al.
2015/0322507 A1 2015/0329891 A1		Zimmermann et al. Tan et al.	2019/0264280			Rabinowitz et al.
2016/0032396 A1		Diehn et al.	2019/0264288			Rabinowitz et al.
2016/0046986 A1		Eltoukhy et al.	2019/0271043 2019/0276888			Babiarz et al. Rabinowitz et al.
2016/0115541 A1 2016/0145682 A1		Schutz et al. Woodward et al.	2019/0284623			Rabinowitz et al.
2016/0186239 A1	6/2016	Sinha	2019/0300950			Rabinowitz et al.
2016/0186253 A1 2016/0201124 A1		Talasaz et al. Donahue et al.	2019/0309358 2019/0309359			Rabinowitz et al. Zimmermann et al.
2016/0239602 A1		Shendure et al.	2019/0309365	A1	10/2019	Babiarz et al.
2016/0244838 A1		Babiarz et al.	2019/0316177 2019/0316184			Zimmermann et al. Zimmermann et al.
2016/0257993 A1 2016/0289740 A1		Fu et al. Fu et al.	2019/0316184		10/2019	
2016/0289753 A1		Osborne et al.	2019/0323076	A1	10/2019	Rabinowitz et al.
2016/0312276 A1		Fu et al.	2019/0360036 2020/0024653			Rabinowitz et al. Bethke
2016/0319345 A1 2016/0369333 A1		Gnerre et al. Babiarz et al.	2020/0024033			Talasaz et al.
2017/0011166 A1		Rabinowitz et al.	2020/0109449			Stamm et al.
2017/0107576 A1		Babiarz et al.	2020/0123612 2020/0126634			Babiarz et al. Sigurjonsson et al.
2017/0114411 A1 2017/0121716 A1	4/2017 5/2017	Mitchell et al. Rodi et al.	2020/0140950			Babiarz et al.
2017/0152561 A1	6/2017	Hamamah et al.	2020/0149111			Babiarz et al.
2017/0218458 A1 2017/0275689 A1		Fan et al.	2020/0157629 2020/0172977			Babiarz et al. Rabinowitz et al.
2017/02/3089 A1 2017/0283788 A1		Maguire et al. Khoja et al.	2020/0181697	$\mathbf{A}1$	6/2020	Rabinowitz et al.
2017/0314014 A1	11/2017		2020/0190570 2020/0190573			Ryan et al. Rabinowitz et al.
2017/0342477 A1 2017/0362649 A1	11/2017	Jensen et al. Lieberman-Aiden et al.	2020/01905/3			Rabinowitz et al.
2018/0023128 A1		Yanai et al.	2020/0208196			Zimmermann et al.
2018/0025109 A1		Rabinowitz et al. Hu et al.	2020/0208221 2020/0224273			Babiarz et al. Rabinowitz et al.
2018/0127744 A1 2018/0148777 A1		Kirkizlar et al.	2020/0232036	A1	7/2020	Rabinowitz et al.
2018/0155775 A1		Zimmermann et al.	2020/0232037 2020/0248264			Babiarz et al. Rabinowitz et al.
2018/0155776 A1 2018/0155779 A1		Zimmermann et al. Zimmermann et al.	2020/0248264			Swanton et al.
2018/0155785 A1		Rabinowitz et al.	2020/0316498			Mitchell
2018/0155786 A1		Rabinowitz et al.	2020/0318191 2020/0350034			Babiarz et al. Rabinowitz et al.
2018/0155792 A1 2018/0171409 A1		Rabinowitz et al. Rabinowitz et al.	2020/0362415			Rabinowitz et al.
2018/0171420 A1		Babiarz et al.	2020/0407788			Ryan et al.
2018/0173845 A1	6/2010	Sigurjonsson et al.	2020/0407798 2021/0009990			Babiarz et al. Stray et al.
2018/0173846 A1 2018/0187241 A1		Sigurjonsson et al. Selvaraj et al.	2021/0025005	A1	1/2021	Babiarz et al.
2018/0201995 A1	7/2018	Rabinowitz et al.	2021/0054459 2021/0071246			Rabinowitz et al. Zimmermann et al.
2018/0237841 A1 2018/0251553 A1	8/2018	Stray et al. McGranahan et al.	2021/00/1240			Mitchell et al.
2018/0265917 A1		Barany et al.	2021/0139983	A1		Mitchell et al.
2018/0288982 A1	10/2018		2021/0139988 2021/0155988			Mitchell et al. Rabinowitz et al.
2018/0298439 A1 2018/0300448 A1		Ryan et al. Rabinowitz et al.	2021/0189498		6/2021	Babiarz et al.
2018/0320171 A1	11/2018	Withey	2021/0198733			Moshkevich et al.
2019/0010543 A1 2019/0106737 A1		Babiarz et al. Underhill	2021/0198742 2021/0198743			Rabinowitz et al. Rabinowitz et al.
2019/0106751 A1		Zimmermann et al.	2021/0222230	A1	7/2021	Zimmermann et al.
2019/0185913 A1		Zimmermann et al.	2021/0222240			Moshkevich et al.
2019/0185936 A1 2019/0194743 A1		Babiarz et al. Ryan et al.	2021/0257048 2021/0269879			Zimmermann B et al. Mitchell et al.
2019/0194743 A1		Babiarz et al.	2021/0324463			Rabinowitz et al.
2019/0194759 A1		Babiarz et al.	2021/0327538			Egilsson et al.
2019/0203290 A1 2019/0203294 A1	7/2019 7/2019		2021/0327542 2021/0355536			Ryan et al. Rabinowitz et al.
2019/0203294 A1 2019/0211391 A1		Rabinowitz et al.	2022/0025455			Zimmermann et al.
2019/0211392 A1	7/2019	Rabinowitz et al.	2022/0025456	A1	1/2022	Rabinowitz et al.

# US 11,519,035 B2

(56)	References Cited	WO	2004087863 A2	10/2004
. ,		WO	2005003375 A2	1/2005
	U.S. PATENT DOCUMENTS	WO WO	2005021793 A1 2005023091 A2	3/2005 3/2005
2022/0	033908 A1 2/2022 Rabinowitz et al.	wo	2005030999 A1	4/2005
	033909 A1 2/2022 Rabinowitz et al.	WO	2005035725 A2	4/2005
2022/0	042103 A1 2/2022 Rabinowitz et al.	WO	2005/039389 A3	5/2005
	056509 A1 2/2022 Zimmermann	WO WO	2005100401 A2 2005123779 A2	10/2005 12/2005
	056534 A1	WO	2003123779 A2 2007145612 A1	6/2006
	073978 A1 3/2022 Rabinowitz et al.	WO	2006110855 A2	10/2006
	073979 A1 3/2022 Rabinowitz et al.	WO	2006/128192 A2	11/2006
		WO WO	2007/011903 A3 2007/052006 A1	1/2007 5/2007
	FOREIGN PATENT DOCUMENTS	wo	2007057647 A1	5/2007
CN	101675169 A 3/2010	WO	2007062164 A3	5/2007
CN	104736722 A 6/2015	WO WO	2007/073171 A2 2007070280 A2	6/2007 6/2007
CN	105229175 A 1/2016	WO	2007070280 A2 2007070482 A2	6/2007
EP	0270017 A2 6/1988	WO	2007075836 A2	7/2007
EP EP	1325963 A1 7/2003 1524321 A1 4/2005	WO	2007/092473 A2	8/2007
EP	1325963 B1 9/2006	WO WO	2007086935 A2 2007/117256 A1	8/2007 10/2007
EP	1524321 B1 7/2009	WO	2007/117230 A1 2007117039 A1	10/2007
EP EP	2163622 A1 3/2010	WO	2007132167 A2	11/2007
EP EP	2128169 A1 12/2010 2902500 A1 8/2015	WO	2007/147073 A2	12/2007
EP	3026124 A1 6/2016	WO WO	2007/147076 A2 2007140417 A2	12/2007 12/2007
EP	3285193 A1 2/2018	wo	2007147074 A2	12/2007
EP	2877594 B1 12/2019 3187597 B1 6/2020	WO	2007147079 A2	12/2007
EP EP	3187597 B1 6/2020 3134541 B1 8/2020	WO	2008024473 A2	2/2008
EP	3760730 A1 1/2021	WO WO	2008048931 A1 2008/061213 A2	4/2008 5/2008
EP	3760731 A1 1/2021	WO	2008051928 A2	5/2008
EP EP	3760732 A1 1/2021 3824470 5/2021	WO	2008056937 A1	5/2008
EP	3443119 B1 2/2022	WO WO	2008059578 A1 2008079374 A2	5/2008 7/2008
GB	2488358 8/2012	WO	2008079374 A2 2008081451 A2	7/2008
JP JP	2965699 8/1999 2002-520121 A 9/2002	WO	2008084405 A2	7/2008
JP JP	2002-530121 A 9/2002 2002/300894 A 10/2002	WO WO	2008115427 A2	9/2008
JР	2003/521252 A 7/2003	WO	2008115497 A2 2008118988 A1	9/2008 10/2008
JР	2004502466 A 1/2004	WO	2008135837 A2	11/2008
JP JP	2004533243 A 11/2004 2005514956 A 5/2005	WO	2008157264 A2	12/2008
JP	2005160470 A 6/2005	WO WO	2009009769 A2 2009013492 A1	1/2009 1/2009
JP	2006-254912 A 9/2006	wo	2009013496 A1	1/2009
JP JP	2008-263974 A 11/2008 2008/271980 A 11/2008	WO	2009019215 A1	2/2009
JP	2010-509922 A 4/2010	WO WO	2009019455 A2 2009/032779 A2	2/2009 3/2009
JP	2011/508662 A 3/2011	WO	2009/036525 A2	3/2009
JР	2011/516069 A 5/2011	WO	2009030100 A1	3/2009
JP RU	2015-535681 12/2015 2290078 C1 12/2006	WO	2009032781 A2	3/2009
WO	95/01796 1/1995	WO WO	2009033178 A1 2009049889 A1	3/2009 4/2009
WO	98/39474 9/1998	wo	2009017784 A2	5/2009
WO WO	98/44151 10/1998 00/18957 4/2000	WO	2009064897 A2	5/2009
wo	0134844 A1 5/2001	WO WO	2009091934 A1 2009092035 A2	7/2009 7/2009
WO	01/57269 A2 8/2001	WO	2009/105531 A1	8/2009
WO	179851 A1 10/2001	WO	2009099602 A1	8/2009
WO WO	200190419 A2 11/2001 2002004672 A2 1/2002	WO	2009100029 A1	8/2009
wo	02/44411 A1 6/2002	WO WO	2009105531 A1 2009117122 A2	8/2009 9/2009
WO	2002055985 A2 7/2002	wo	2009120808 A2	10/2009
WO WO	02/070751 A1 9/2002 2002076377 10/2002	WO	2009145828 A2	12/2009
WO	02/090505 A2 11/2002	WO	2009146335 A1	12/2009
WO	03/000919 A2 1/2003	WO WO	2010014920 A1 2010017214 A1	2/2010 2/2010
WO	03/018757 A3 3/2003	WO	2010033578 A1	2/2010
WO WO	03/020974 A3 3/2003 2003031646 A1 4/2003	WO	2010/033639 A2	3/2010
WO	3050532 A1 6/2003	WO	2010/033652 A1	3/2010
WO	2003062441 A1 7/2003	WO WO	2010042831 A2 2010045617 A2	4/2010 4/2010
WO	3102595 A1 12/2003	WO	2010043617 A2 2010075459	7/2010
WO WO	3106623 A2 12/2003 2004/051218 A2 6/2004	wo	2010/088288 A2	8/2010
wo	2004/051218 A2 0/2004 2004069849 A2 8/2004	WO	2010115016 A2	10/2010
WO	2004070005 A2 8/2004	WO	2010115154 A1	10/2010
WO	2004070007 A2 8/2004	WO	2010118016 A2	10/2010

# US 11,519,035 B2

(56)	References Cited	WO 2018/237081 A1 12/2018
· /	FOREIGN PATENT DOCUMENTS	WO 2019/046817 A1 3/2019 WO 2019/118926 A1 6/2019
		WO 2019/140298 A1 7/2019
WO WO	2010/127186 A1 11/2010 2011/023078 A1 3/2011	WO 2019/161244 A1 8/2019 WO 2019/200228 A1 10/2019
WO	2011/023078 A1 3/2011 2011/032078 A1 3/2011	WO 2019/241349 A1 12/2019
WO	2011041485 A1 4/2011	WO 2020/010255 A1 1/2020 WO 2020/018522 A1 1/2020
WO WO	2011/051283 A1 5/2011 2011057061 A1 5/2011	WO 2020/041449 A1 2/2020
WO	2011057094 5/2011	WO 2020/076957 A1 4/2020
WO WO	2011/090556 A1 7/2011 2011087760 7/2011	WO 2020/106987 A1 5/2020 WO 2020/131699 A2 6/2020
wo	2011037700 7/2011 2011102998 A2 8/2011	WO 2020/214547 A1 10/2020
WO	2011/118603 9/2011	WO 2020/247263 A1 12/2020 WO 2021/055968 A1 3/2021
WO WO	2011109440 A1 9/2011 2011140433 A2 11/2011	WO 2007100911 A2 9/2021
WO	2011146632 A1 11/2011	WO 2021/243045 A1 12/2021 WO 2022/015676 A1 1/2022
WO WO	2012/019200 A2 2/2012 2012/028746 A1 3/2012	WO 2022/013070 A1 1/2022
WO	2012042374 A2 4/2012	OTHER PUBLICATIONS
WO WO	2012/058488 A1 5/2012 201283250 6/2012	OTTEN TOBERCHITORIO
wo	2012088456 A2 6/2012	Wright et al. (Hum Reprod Updt, 2009, 15(1):139-151) (Year:
WO	20120071621 6/2012	2009).*  Cronn et al. (Nucleic Acids Research 2008, 36(19):e122) (Year:
WO WO	2012108920 A1 8/2012 2012/142531 A2 10/2012	2008).*
WO	2007/149791 A2 12/2012	Su et al. (J of Molec Diagn, 2011, 13(1):74-84) (Year: 2011).*
WO WO	2013030577 3/2013 2013/045432 A1 4/2013	Bashashati, A. et al., "Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial muta-
WO	2013/049892 A1 4/2013	tional profiling", Journal of Pathology, vol. 231, 2013, 21-34.
WO WO	2013052557 A2 4/2013 2013/078470 A2 5/2013	Binladen, J. et al., "The Use of Coded PCR Primers Enables
wo	2013/086464 A1 6/2013	High-Throughput Sequencing of Multiple Homolog Amplification
WO	2013/123220 A1 8/2013 2013/138510 A1 0/2013	Products by 454 Parallel Sequencing", PLOS One, Issue 2, Feb. 2007, 9 pages.
WO WO	2013/138510 A1 9/2013 2013/138510 A9 9/2013	De Jong, M. M. et al., "Genes other than BRCA 1 and BRCA2
WO	20130130848 9/2013	involved in breast cancer susceptibility", J. Med. Genet., vol. 39,
WO WO	2013/159035 A2 10/2013 2013/169339 A1 11/2013	2009, 225-242. Faham, M. et al., "Deep-sequencing approach for minimal residual
WO	2013/177220 A1 11/2013	disease detection in acute lymphoblastic leukemia", Blood Journal,
WO WO	2013/181651 A1 12/2013 2014/004726 A1 1/2014	vol. 120, No. 26, Dec. 20, 2012, 5173-5180.
WO	2014/014497 A1 1/2014	Gao, F. et al., "Characterizing Immunoglobulin Repertoire from Whole Blood by a Personal Genome Sequencer", PLOS One, vol.
WO WO	20140018080 1/2014 2014/035986 A1 3/2014	8, No. 9, Sep. 13, 2013, 8 pgs.
WO	2014/122288 A1 8/2014	Gnirke, A. et al., "Solution hybrid selection with ultra-long oligo-
WO WO	2014/145078 A1 9/2014 2014/145232 A2 9/2014	nucleotides for massively parallel targeted sequencing", Nature Biotechnology, vol. 27, No. 2, Feb. 2009, 182-189.
wo	2014/149134 A2 9/2014	Hodgkinson, C. L. et al., "Tumorigenicity and genetic profiling of
WO WO	2014/150300 A2 9/2014	circulating tumor cells in small-cell lung cancer", Nature Medicine,
WO	2014/151117 A1 9/2014 2014/194113 A2 12/2014	vol. 20, No. 8, Aug. 2014, 897-905. Hou, X. et al., "Analysis of the Repertoire Features of TCR Beta
WO	2015134552 A1 3/2015	Chain CDR3 in Human by High-Throughput Sequencing", Cellular
WO WO	2015/048535 A1 4/2015 2015/100427 A1 7/2015	Physiology and Biochemistry, vol. 39, Jul. 21, 2019, 651-667.
WO	2015/148494 A1 10/2015	Illumina, "MiSeq System Information Sheet", 2018, 3 pgs. Keller, M. C. et al., "Non-Pathological Paternal Isodisomy of
WO WO	2015/164432 A1 10/2015 2016/009059 A1 1/2016	Chromosome 2 Detected From a Genome-Wide SNP Scan", Ameri-
WO	2016009224 A1 1/2016	can Journal of Medical Genetics, Part A, 2009, 1823-1826.
WO WO	2016/063122 A1 4/2016 2016/065295 A1 4/2016	Kukita, Y. et al., "High-fidelity target sequencing of individual molecules identified using barcode sequences: de nova detection
wo	2016/077313 A1 5/2016	and absolute quantitation of mutations in plasma cell-free DNA
WO WO	2016/123698 A1 8/2016 2016/138080 A1 9/2016	from cancer patients", DNA Research, vol. 22, No. 4, Jun. 29, 2015,
WO	2016/176662 A1 11/2016	269-277. Scarpa, A. et al., "Molecular Typing of Lung Adenocarcinoma on
WO	2016/183106 A1 11/2016	Cytological Samples Using a Multigene Next Generation Sequenc-
WO WO	2016/193490 A1 12/2016 2017/058784 A1 4/2017	ing Panel", PLOS One, vol. 8, No. 11, Nov. 13, 2013, 6 pgs.
WO	2017/181146 A1 10/2017	Short, N. J. et al., "Targeted next-generation sequencing of circulating cell-free DNA vs bone marrow in patients with acute myeloid
WO WO	2017/181202 A2 10/2017 2017/190106 A1 11/2017	leukemia", Blood Advances, vol. 4, No. 8, Apr. 23, 2020, 1670-
WO	2017205540 A1 11/2017	1677.
WO WO	2018/009723 A1 1/2018 2018/083467 A1 5/2018	Siebert, P. D. et al., "An improved PCR method for walking in uncloned genomic DNA", Nucleic Acids Research, vol. 23, No. 6,
WO	2018/085407 A1	1995, 1087-1088.
WO	2018/106798 A1 6/2018	Tewhey, R. et al., "Microdroplet-based PCR enrichment for large-
WO WO	2018/136562 A2 7/2018 2018/156418 A1 8/2018	scale targeted sequencing", Nature Biotechnology, vol. 27, No. 11, Nov. 2009, 1025-1031.
9		

Case: 24-1324 Document: 42-1 Page: 43 Filed: 03/18/2024

### US 11,519,035 B2

Page 8

#### References Cited (56)

#### OTHER PUBLICATIONS

Zhou, W. et al., "Counting alleles to predict recurrence of earlystage colorectal cancers", The Lancet, vol. 359, Jan. 19, 2002,

"European Application No. 014198110, European Search Report dated Apr. 28, 2015, 3 pages.

"Finishing the Euchromatic Sequence of the Human Genome", Nature vol. 431,(Oct. 21, 2004),931-945.

"FixedMedium, dictionary definition, Academic Press Dictionary of Science and Technology", Retrieved from the Internet: <URL:www. credoreference.com/entry/apdst/fixed\_medium>, 1996, 1 pg.

"GeneticsHome Reference", http://ghr.nlm.nih.gov/handbook/ genomicresearch/snp, Feb. 28, 2014, 1-2.

"Ion Ampli Seq Comprehensive Cancer Panel, product brochure, Life TechnologiesCorporation. Retrieved from the Internet", <URL:https://tools.lifetechnologies.com/content/sfs/brochures/Ion\_</p> Comp CancerPanel\_Flyer.pdf>, 2012, 2 pgs

"Multiplexing with RainDrop Digital PCR", RainDance Technologies, Application Note, 2013, 1-2.

NucleicAcids, Linkers and Primers: Random Primers", New England BioLabs 1998/99Catalog, 1998, 121 and 284.

"PRIMER3, information sheet, Sourceforge.net. [retrieved on Nov. 12, 2012]. Retrieved from the Internet: <URL: http://primer3. sourceforge.net/>", 2009, 1 pg.

"Db SNP rs2056688 (http://www.ncbi.nlm.nih.gov/projects/SNP/ snp\_ref.cgi?rs=2056688, downloaded May 4, 2015", 3 pages

PRNewswire (Research Suggests Daily Consumption of Orange Juice Can Reduce Blood Pressure and May Provide Beneficial Effects to Blood Vessel Function: New Study Identified Health Benefits in Orange Juice, Dec. 8, 2010), 3 pages

What to Expect (Weird Harmony results, May 1, 2015), 7 pages. The Bump (Panorama Test, attached, Jul. 1, 2013), 8 pages

"Blast of AAAAAAAATTTAAAAAAAATTT(http://blast.ncbi. nlm.nih.gov/Blast.cgi, downloaded May 4, 2015)", 9 pages

"Guideline related to genetic examination", Societies Related to Genetic Medicine, Japanese Society for Genetic Counseling, Japanese Society for Gene Diagnosis and Therapy, Japan Society of Obstetrics and Gynecology, 2003, 2-15.

"How Many Carbs in a Potato?, [Online]", Retrieved from the Internet: <a href="http://www.newhealthguide.org/How-Many-Carbs-In-A-">http://www.newhealthguide.org/How-Many-Carbs-In-A-</a> Potato.html>, Nov. 1, 2014, 3 pages.

"Random variable", In The Penguin Dictionary of Mathematics. Retrieved from http://www.credoreference.com/entry/penguinmath/

random\_variable, 2008, 1 page. Abbosh, C. et al., "Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution", Nature, vol. 545, May 25, 2017, 446-451. Abidi, S. et al., "Leveraging XML-based electronic medical records to extract experiential clinical knowledge: An automated approach to generate cases for medical case-based reasoning systems", International Journal of Medical Informatics, 68(1-3), 2002, 187-203. Agarwal, Ashwin. et al., "Commercial Landscape of Noninvasive Prenatal Testing in the United States", Prenatal Diagnosis, 33, 2013,

521-531 Alaeddini, R. et al., "Forensic implications of genetic analyses from degraded DNA—A review", Forensic Science International: Genet-

ics, vol. 4, 2010, 148-157. Alberts, B. et al., "Chapter 20: Germ Cells and Fertilization", Molecular Biology of the Cell, Fourth Edition, 2002, 1127-1156. Alberts, B. et al., "Chapter 4: DNA and Chromosomes", Molecular

Biology of the Cell, Fourth Edition, 2002, 191-234. Alkan, Can et al., "Personalized Copy Number and Segmental Duplication Maps Using Next-Generation Sequencing", Nature Genetics, 41, 10, 2009, 1061-1068.

Allaire, F R., "Mate selection by selection index theory", Theoretical Applied Genetics, 57(6), 1980, 267-272.

Allan, J. et al., "Micrococcal Nuclease Does Not Substantially Bias Nucleosome Mapping", Journal of Molecular Biology, vol. 417, Jan. 30, 2012, 152-164.

Allawi, Hatim T. et al., "Thermodynamics of internal C•T Mismatches in DNA", Nucleic Acids Research, 26 (11), 1998, 2694-

Andras, S. C. et al., "Strategies for Signal Amplification in Nucleic Acid Detection", Molecular Biotechnology, vol. 19, 2001, 29-44. Anker, P. et al., "Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients", Cancer and Metastasis Reviews, vol. 18, 1999, 65-73.

Antonarakis, S. E. et al., "Chromosome 21 and Down Syndrome: From Genomics to Pathophysiology", Nature Reviews Genetics, vol. 5, Oct. 2004, 725-738.

Aoki, Yasuhiro, "Statistical and Probabilistic Bases of Forensic DNA Testing", The Journal of the Iwate Medical Association, 2002, vol. 54, p. 81-94.

Ashoor, G. et al., "Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics", Ultrasound in Obstetrics and Gynecology, vol. 41, 2013,

Ashoor, Ghalia et al., "Chromosome-Selective Sequencing of Maternal Plasma Cell-Free DNA for First-Trimester Detection of Trisomy 21 and Trisomy 18", American Journal of Obstetrics & Gynecology, 206, 2012, 322.e1-322.e5.

Ashoor, Ghalia et al., "Fetal Fraction in Maternal Plasma Cell-Free DNA at 11-13 Weeks' Gestation: Effect of Maternal and Fetal Factors", Fetal Diagnosis Therapy, 2012, 1-7.

Bada, Michael A. et al., "Computational Modeling of Structural Experimental Data", Methods in Enzymology, 317, 2000, 470-491. Bai, H. et al., "Detection and Clinical Significance of Intratumoral EGFR Mutational Heterogeneity in Chinese Patients with Advanced Non-Small Cell Lung Cancer", PLOS One, vol. 8, No. 2, Feb. 2013, 7 pages

Ballif, B. C. et al., "Detection of Low-Level Mosaicism by Array CGH in Routine Diagnostic Specimens", American Journal of Medical Genetics Part A, vol. 140A, 2006, 2757-2767.

Beaumont, Mark A et al., "The Bayesian Revolution in Genetics", Nature Reviews Genetics, 5, 2004, 251-261.

Beck, J. et al., "Digital Droplet PCR for Rapid Quantification of Donor DNA in the Circulation of Transplant Recipients as a Potential Universal Biomarker of Graft Injury", Clinical Chemistry,

vol. 59, No. 12, 2013, 1732-1741. Beer, Alan E. et al., "The Biological Basis of Passage of Fetal Cellular Material into the Maternal Circulation", Annals New York Academy of Sciences, 731, 1994, 21-35. Beerenwinkel, et al., "Methods for Optimizing Antiviral Combina-

tion Therapies", Bioinformatics, 19(1), 2003, i16-i25.

Beerenwinkel, N. et al., "Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes", Nucleic Acids Research, 31(13), 2003, 3850-3855.

Benn, P. et al., "Non-Invasive Prenatal Testing for Aneuploidy: Current Status and Future Prospects", Ultrasound Obstet Gynecol, 42, 2013, 15-33.

Benn, P et al., "Non-Invasive prenatal Diagnosis for Down Syndrome: the Paradigm Will Shift, but Slowly", Ultrasound Obstet. Gynecol., 39, 2012, 127-130,

Bentley, David R et al., "Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry", Nature, 456, 6, 2008,

Bermudez, M. et al., "Single-cell sequencing and mini-sequencing for preimplantation genetic diagnosis", Prenatal Diagnosis, 23, 2003, 669-677.

Beroud, C. et al., "Prenatal diagnosis of spinal muscular atrophy by genetic analysis of circulating fetal cells", The Lancet, vol. 361, Mar. 22, 2003, 1013-1014.

Bevinetto, Gina, Bevinetto (5 Foods All Pregnant Women Need, American Baby, available at http://www.parents.com/pregnancy/ mybody/nutrition/5greatpregnancyfoods/, Apr. 15, 2008), 8 pgs.

Bianchi, D W. et al., "Fetal gender and aneuploidy detection using fetal cells maternal blood: analysis of NIFTY I data", Prenat Diagn 2002; 22, 2002, 609-615

Bianchi, D. W., "Circulating Fetal DNA: Its Origin and Diagnostic Potential-A Review", Placenta, vol. 25, Supplemental A, May 2004, S93-S101.

### US 11,519,035 B2

Page 9

#### (56) References Cited

#### OTHER PUBLICATIONS

Bianchi, D. W. et al., "Genome-Wide Fetal Aneuploidy Detection by Maternal Plasma DNA Sequencing", Obstetrics & Gynecology, vol. 119, No. 5, May 2012, 890-901.

Bianchi, D. W., "Review: Fetal Cells in the Maternal Circulation: Feasibility for Prenatal Diagnosis", British Journal of Haematology, vol. 105, 1999, 574-583.

Birch, Lyndsey et al., "Accurate and Robust Quantification of Circulating Fetal and Total DNA in Maternal Plasma from 5 to 41 Weeks of Gestation", Clinical Chemistry, 51(2), 2005, 312-320.

Birkenkamp-Demtroder, K. et al., "Abstract 3653: Sequencing of plasma cfDNA from patients with locally advanced bladder cancer for surveillance and therapeutic efficacy monitoring", Cancer Research, vol. 78, No. 13 Supplement, Jul. 2019, 1 page.

Bisignano, et al., "PGD and Aneuploidy Screening for 24 Chromosomes: Advantages and Disadvantages of Competing Platforms", Reproductive BioMedicine Online, 23, 2011, 677-685.

Bodenreider, O., "The Unified Medical Language System (UMLS): Integrating Biomedical Terminology", Nucleic Acids Research, 32, (Database issue), 2004, D267-D270.

Breithaupt, Holger, "The Future of Medicine", EMBO Reports, 21(61), 2001, 465-467.

Brownie, Jannine et al., "The Elimination of Primer-Dimer Accumulation in PCR", Nucleic Acids Research, 25(16), 1997, 3235-3241.

Burkova, E. E. et al., "Extremely Stable Soluble High Molecular Mass Multi-Protein Complex with DNase Activity in Human Placental Tissue", PLOS One, vol. 9, No. 11: e011234, Nov. 26, 2014, 26 pages.

Burnham, P. et al., "Myriad Applications of Circulating Cell-Free DNA in Precision Organ Transplant Monitoring", Annals of the American Thoracic Society, vol. 14, Supplement 3, Sep. 2017, S237-S241.

Butler, J. et al., "The Development of Reduced Size STR Amplicons as Tools for Analysis of Degraded DNA", Journal of Forensic Sciences, vol. 48, No. 5, 2003, 1054-1064.

Butt, A. N. et al., "Overview of Circulating Nucleic Acids in Plasma/Serum: Update on Potential Prognostic and Diagnostic Value in Diseases Excluding Fetal Medicine and Oncology", Ann. N.Y. Acad. Sci., vol. 1137, 2008, 236-242.

Cairns, Paul et al., "Homozygous Deletions of 9p21 in Primary Human Bladder Tumors Detected by Comparative Multiplex Polymerase Chain Reaction", Cancer Research, 54, 1994, 1422-1424. Caliendo, Angela, "Multiplex PCR and Emerging Technologies for the Detection of Respiratory Pathogens", Clinical Infectious Diseases, 52(4), 2011, 8326-8330.

eases, 52(4), 2011, S326-S330.

Cansar, ,"Hs-578-T—Copy Number Variation—Cell Line Synopsis", ICR Cancer Research UK, Retrieved on Mar. 26, 2018 from https://cansari.cr.ac.uk/cansar/cell-lines/Hs-578-T/copy\_number\_variation/chromosome 8/. Mar. 26, 2018, 50 pgs.

tion/chromosome\_8/, Mar. 26, 2018, 50 pgs. Carnevale, Alessandra et al., "Attitudes of Mexican Geneticists Towards Prenatal Diagnosis and Selective Abortion", American Journal of Medical Genetics, 75, 1998, 426-431.

Carvalho, B. et al., "Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data", Biostatistics, vol. 8, No. 2, 2007, 485-499.

Casbon, J. A. et al., "A method for counting PCR template molecules with application to next-generation sequencing", Nucleic Acids Research, vol. 39, No. 12, Apr. 13, 2011, 1-8. Chakraborty, R. et al., "Paternity Exclusion by DNA Markers:

Chakraborty, R. et al., "Paternity Exclusion by DNA Markers: Effects of Paternal Mutations", Journal of Forensic Sciences, vol. 41, No. 4, Jul. 1996, 671-677.

Chan, K.C. et al., "Size Distributions of Maternal and Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 50, No. 1, 2004, 88-92. Chang, H.W. et al., "Assessment of Plasma DNA Levels, Allelic Imbalance, and CA 125 as Diagnostic Tests for Cancer", Journal of the National Cancer Institute, vol. 94, No. 22, Nov. 20, 2002, 1697-1703.

Chen, E. et al., "Noninvasive Prenatal Diagnosis of Fetal Trisomy 18 and Trisomy 13 by Maternal Plasma DNA Sequencing", PLoS ONE, 6 (7), e21791, 2011, 7 pgs.

Chen, X. Q. et al., "Microsatallite alterations in plasma DNA of small cell lung cancer patients", Nature Medicine, vol. 2, No. 9, Sep. 1996, 1033-1035.

Chetty, Shilpa et al., "Uptake of Noninvasive Prenatal Testing (NIPT) in Women Following Positive Aneuploidy Screening", Prenatal Diagnosis, 33, 2013, 542-546.

Cheung, S. W. et al., "Rapid Publication: Microarray-Based CGH Detects Chromosomal Mosaicism Not Revealed by Conventional Cytogenetics", American Journal of Medical Genetics Part A, vol. 143A, 2007, 1679-1686.

Cheung, V. G. et al., "Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA", Proceedings of the National Academy of Sciences, USA, vol. 93, Dec. 1996, 14676-14679.

Chiu, R. et al., "Non-Invasive Prenatal Assessment of Trisomy 21 by Multiplexed Maternal Plasma DNA Sequencing: Large Scale Validity Study", BMJ, 342, c7401, 2011, 9 pgs. Chiu, Rossa W. et al., "Effects of Blood-Processing Protocols on

Chiu, Rossa W. et al., "Effects of Blood-Processing Protocols on Fetal and Total DNA Quantification in Maternal Plasma", Clinical Chemistry, 47(9), 2001, 1607-1613.

Chiu, Rossa W.K. et al., "Maternal Plasma DNA Analysis with Massively Parallel Sequencing by Litigation for Noninvasive Prenatal Diagnosis of Trisomy 21", Clinical Chemistry, 56, 3, 2010, 459-463.

Chiu, Rossa W.K. et al., "Non-Invasive Prenatal Diagnosis by Single Molecule Counting Technologies", Trends in Genetics, 25 (7), 2009, 324-331.

Chiu, Rossa W.K. et al., "Noninvasive Prenatal Diagnosis of Fetal Chromosomal Aneuploidy by Massively Parallel Genomic Sequencing of DNA in Maternal Plasma (with Supporting Information)", PNAS, vol. 105, No. 51, 2008, 20458-20463.

Choi, M. et al., "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing", PNAS, vol. 106, No. 45, Nov. 10, 2009, 19096-19101.

Chu, T. et al., "Statistical Considerations for Digital Approaches to Non-Invasive Fetal Genotyping", Bioinformatics (Advance Access publication), 26 (22), 2010, 2863-2866.

Chu, Tianjiao et al., "Statistical Model for Whole Genome Sequencing and its Application to Minimally Invasive Diagnosis of Fetal Genetic Disease", Bioinformatics, 25(10), 2009, 1244-1250. Chu, Tianjiao. et al., "A Novel Approach Toward the Challenge of

Chu, Tianjiao. et al., "A Novel Approach Toward the Challenge of Accurately Quantifying Fetal DNA in Maternal Plasma", Prenatal Diagnosis, 30, 2010, 1226-1229.

Cole, Neal W. et al., "Hyperglycemia-Induced Membrane Lipid Peroxidation and Elevated Homocysteine Levels Are Poorly Attenuated by Exogenous Folate in Embryonic Chick Brains", Comparative Biochemistry and Physiology, Part B, 150, 2008, 338-343.

Colella, S. et al., "QuantiSNP: an Objectives Bayes Hidden-Markov Model to Detect and Accurately Map Copy Number Variation Using SNP Genotyping Data", Nucleic Acids Research, 35 (6), 2007, 2013-2025.

Conlin, L. K. et al., "Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis", Human Molecular Genetics, vol. 19, No. 7, Jan. 6, 2010, 1263-1275.

Coombes, R. C. , "Abstract P4-01-02: Early detection of residual breast cancer through a robust, scalable and personalized analysis of circulating tumour DNA (ctDNA) antedates overt metastatic recurrence", Cancer Research, vol. 79, No. 4 Supplement, Feb. 15, 2019. Cossu, Gianfranco et al., "Rh D/d Genotyping by Quantitative Polymerase Chain Reaction and Capillary Zone Electrophoresis", Electrophoresis, 17, 1996, 1911-1915.

Coyle, J. F. et al., "Standards for detailed clinical models as the basis for medical data exchange and decision support", International Journal of Medical Informatics, 69(2-3), 2003, 157-174.

Craig, D. W. et al., "Identification of genetic variants using barcoded multiplexed sequencing", Nature Methods, vol. 5, Oct. 2008, 887-893.

### US 11,519,035 B2

Page 10

#### (56) References Cited

#### OTHER PUBLICATIONS

Cross, Jillian et al., "Resolution of trisomic mosaicism in prenatal diagnosis: estimated performance of a 50K SNP microarray", Prenat Diagn 2007; 27, 2007, 1197-1204.

D'Aquila, Richard et al., "Maximizing sensitivity and specificity of PCR by pre-amplification heating", Nucleic Acids Research, 19(13), 1991, p. 3749.

Daruwala, Raoul-Sam et al., "A Versatile Statistical Analysis Algorithm to Detect Genome Copy Number Variation", PNAS, 101(46), 2004, 16292-16297.

Dawson, S.J. et al., "Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer", The New England Journal of Medicine, vol. 368, No. 13, Mar. 28, 20136, 1199-1209.

De Bruin, E. et al., "Spatial and temporal diversity in genomic instability processes defines lung cancer evolution", Science, vol. 346, No. 6206, Oct. 10, 2014, 251-256.

De Vries, et al., "Diagnostic genome profiling in mental retardation", Am J Hum Genet, 77, published online Aug. 30, 2005, 2005, 606-616

Deangelis, M. et al., "Solid-phase Reversible Immobilization for the Isolation of PCR Products", Nucleic Acids Research, 23 (22), 1995, 4742-4743

Deng, S. et al., "TNER: A Novel Background Error Suppression Method for Mutation Detection in Circulating Tumor DNA", bioRxiv, http://dx.doi.org/10.1101/214379, Nov. 5, 2017, 12 pgs.

Deutsch, S. et al., "Detection of aneuploidies by paralogous sequence quantification", J Med Genet, vol. 41, 2004, 908-915.

Devaney, S. et al., "Noninvasive Fetal Sex Determination Using Cell-Free Fetal DNA: A Systematic Review and Meta-analysis", JAMA, 306 (6), 2011, 627-636.

Dhallan, et al., "Methods to Increase the Percentage of Free Fetal DNA Recovered from the Maternal Circulation", JAMA, 291(9), 2004, 1114-1119.

Dhallan, Ravinder et al., "A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study", The Lancet, 369, 2007, 474-481.

Dieffenbach, C W. et al., "General concepts for PCR primer design", Genome Research. PCR methods and Applications vol. 3, 1993, S30-S37.

Diehl, F. et al., "Circulating mutant DNA to assess tumor dynamics", Nature Medicine, vol. 14, No. 9, Jul. 31, 2008, 985-990.

Diehl, F. et al., "Detection and quantification of mutations in the plasma of patients with colorectal tumors", PNAS, vol. 102, No. 45, Nov. 8, 2005, 16368-16373

Nov. 8, 2005, 16368-16373. Dietmaier, W. et al., "Multiple Mutation Analyses in Single Tumor Cells with Improved Whole Genome Amplification", American Journal of Pathology, vol. 154, No. 1, Jan. 1999, 83-95.

Ding, C et al., "Direct molecular haplotyping of long-range genomic DNA with M1-PCR", PNAS 100(13), 2003, 7449-7453. Dodge, Y. , "Bayes' Theorem", The Concise Encyclopedia of

Dodge, Y., "Bayes' Theorem", The Concise Encyclopedia of Statistics, 2008, 30-31.

Dohm, J. et al., "Substantial Biases in Ultra-Short Read Data Sets

Dohm, J. et al., "Substantial Biases in Ultra-Short Read Data Sets From High-Throughput DNA Sequencing", Nucleic Acids Research, 36 (16), e105, 2008, 10 pag.

36 (16), e105, 2008, 10 pgs.

Dolganov, Gregory et al., "A Novel Method of Gene Transcript Profiling in Airway Biopsy Homogenates Reveals Increased Expression of a Na—K+—Cl-Cotransporter (NKCC1) in Asthmatic Subjects", Genome Res.,11, 2001, 1473-1483.

Donaghue, C. et al., "Detection of mosaicism for primary trisomies in prenatal samples by QF-PCR and karyotype analysis", Prenatal Diagnosis, vol. 25, 2005, 65-72.

Donohoe, Gerard G et al., "Rapid Single-Tube Screening of the C282Y Hemochromatosis Mutation by Real-Time Multiplex Allelespecific PCR without Fluorescent Probes", Clinical Chemistry, 46, 10, 2000, 1540-1547.

Donoso, P. et al., "Current Value of Preimplantation Genetic Aneuploidy Screening in IVF", Human Reproduction Update, 13(1), 2007, 15-25.

Echeverri, et al., "Caffeine's Vascular Mechanisms of Action", International Journal of Vascular Medicine vol. 2010(2010), 10 pages, Aug. 25, 2010.

Ehrich, Mathias et al., "Noninvasive Detection of Fetal Trisomy 21 by Sequencing of DNA in Maternal Blood: A Study in a Clinical Setting", American Journal of Obstetrics & Gynecology, 204, 2011, 205.e1-205.e11.

Eichler, H, "Mild Course of Fetal Rh D Haemolytic Disease due to Maternal Alloimmunisation to Paternal HLA Class I and II Antigens", Vox Sang, 68, 1995, 243-247.

Ellison, Aaron M., "Bayesian Inference in Ecology", Ecology Letters, vol. 7, 2004, 509-520.

Ellonen, P. et al., "Development of SNP Microarray for Supplementary Paternity Testing", International Congress Series, 1261, 2004, 12-14.

EP06838311.6, , "European Communication and Extended European Search Report", dated Dec.  $30,\,2008,\,8$  pgs.

EP08742125.1, , "European Communication pursuant to Article 94(3) EPC and Examination Report", dated Feb. 12, 2010, 5 pgs. Everitt, B. S., "Medical Statistics From A to Z", 2003, 3 pages. Fan, et al., "Whole-genome molecular haplotyping of single cells", Nature Biotechnology, vol. 29, No. 1, Jan. 1, 2011, 51-57. Fan, Christina H. et al., "Non-Invasive Prenatal Measurement of the

Fan, Christina H. et al., "Non-Invasive Prenatal Measurement of the Fetal Genome", Nature, doi:10.1038/nature11251, 2012, 26 pgs. Fan, Christina H et al., "Noninvasive Diagnosis of Fetal Aneuploidy by Shotgun Sequencing DNA from Maternal Blood", PNAS, 105, 42, 2008, 16266-16271.

Fan, H. C. et al., "Microfluidic digital PCR enables rapid prenatal diagnosis of fetal aneuploidy", American Journal of Obstetrics & Gynecology, vol. 200, May 2009, 543.e1-543.e7.
Fan, H. Christina et al., "Sensitivity of Noninvasive Prenatal

Fan, H. Christina et al., "Sensitivity of Noninvasive Prenatal Detection of Fetal Aneuploidy from Maternal Plasma Using Shotgun Sequencing Is Limited Only by Counting Statistics", PLoS One, vol. 5, Issue 5 (e10439), May 3, 2010, 1-6.

Fan, Jian-Bing et al., "Highly Parallel Genomic Assay", Nature Reviews, 7, 2006, 632-644.

Fat Secret, , "5 Foods to Never Eat", "www.fatsecret.com" (printed from internet Nov. 1, 2014)., 2 pages.

Fazio, Gennaro. et al., "Identification of Rapd Markers Linked to Fusarium Crown and Root Rot Resistance (Frl) in Tomato", Euphytica 105, 1999, 205-210.

Fiorentino, F. et al., "Development and Clinical Application of a Strategy for Preimplantation Genetic Diagnosis of Single Gene Disorders Combined with HLA Matching", Molecular Human Reproduction (Advance Access publication), 10 (6), 2004, 445-460. Fiorentino, F et al., "Strategies and Clinical Outcome of 250 Cycles of Preimplantation Genetic Diagnosis for Single Gene Disorders", Human Reproduction, 21, 3, 2006, 670-684.

Human Reproduction, 21, 3, 2006, 670-684.
Fiorentino, Francesco et al., "Short Tandem Repeats Haplotyping of the HLA Region in Preimplantation HLA Matching", European Journal of Human Genetics, 13, 2005, 953-958.
Ford, E. et al., "A method for generating highly multiplexed

Ford, E. et al., "A method for generating highly multiplexed ChIP-seq libraries", BMC Research Notes, vol. 7, No. 312, May 22, 2014, 1-5.

Forejt, et al., "Segmental trisomy of mouse chromosome 17: introducing an alternative model of Down's syndrome", Genomics, 4(6), 2003, 647-652.

Forshew, et al., "Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA", Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci. Transl. Med. 4, 136 30 (2012)., 1-12.

Forshew, T. et al., "Supplementary Materials for Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA", Sci. Transl. Med, vol. 4, May 30, 2012, 20 pgs.

Fouquet, C. et al., "Rapid and Sensitive p53 Alteration Analysis in Biopsies from Lung Cancer Patients Using a Functional Assay and a Universal Oligonudeotide Array: A Prospective Study", Clinical Cancer Research, vol. 10, May 15, 2004, 3479-3489.

Fredriksson, et al., "Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector", Nucleic Acids Research, 2007, vol. 35, No. 7 e47, 1-6.

## US 11,519,035 B2

Page 11

### (56) References Cited

#### OTHER PUBLICATIONS

Freeman, Jennifer L. et al., "Copy Number Variation: New Insights in Genome Diversity", Genome Research, 16, 2006, 949-961. Frost, Mackenzie S et al., "Differential Effects of Chronic Pulsatile Versus Chronic Constant Maternal Hyperglycemia on Fetal Pancreatic B-Cells", Journal of Pregnancy, 2012 Article ID 812094, 2012,

Fu, G. K. et al., "Counting individual DNA molecules by the stochastic attachment of diverse labels", PNAS, vol. 108, No. 22, May 31, 2011, 9026-9031.

Fu, G. K. et al., "Digital Encoding of Cellular mRNAs Enabling Precise and Absolute Gene Expression Measurement by Single-Molecule Counting", Analytical Chemistry, vol. 86, Mar. 3, 2014, 2867-2870.

Ganshirt-Ahlert, D. et al., "Ratio of Fetal to Maternal DNA is Less Than 1 in 5000 at different Gestational Ages in Maternal Blood", Clinical Genetics, 38, 1990, 38-43.

Ganshirt-Ahlert, D. et al., "Fetal DNA in Uterine Vein Blood", Obstetrics & Gynecology, 80 (4), 1992, 601-603.

Ganshirt-Ahlert, Dorothee et al., "Three Cases of 45,X/46,XYnf Mosaicism", Human Genetics, 76, 1987, 153-156.

Garcia-Murillas, I. et al., "Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer", Science Translational Medicine, vol. 7, No. 302, Aug. 26, 2015, 1-2. Gardina, P. et al., "Ploidy Status and Copy Number Aberrations in

Gardina, P. et al., "Ploidy Status and Copy Number Aberrations in Primary Glioblastomas Defined by Integrated Analysis of Allelic Ratios, Signal Ratios and Loss of Heterozygosity Using 500K SNP Mapping Arrays", BMC Genomics, 9 (489), (doi:10.1186/1471-2164-9-489), 2008, 16 pgs.

Geiss, G. K. et al., "Direct multiplexed measurement of gene expression with color-coded probe pairs", Nature Biotechnology, vol. 26, No. 3, Feb. 17, 2008, 317-325.

Ghanta, Sujana et al., "Non-Invasive Prenatal Detection of Trisomy 21 Using Tandem Single Nucleotide Polymorphisms", PLoS ONE, 5 (10), 2010, 10 pgs.

Gholami, M. et al., "A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants", Plant Biotechnology Journal, vol. 10, 2012, 635-645.

Gielis, E. M. et al., "Plasma donor-derived cell-free DNA kinetics after kidney transplantation using a single tube multiplex PCR assay", PLOS One, vol. 13, No. 12, e0208207, Dec. 6, 2018, 16 pgs. Gjertson, David W. et al., "Assessing Probability of Paternity and the Product Rule in DNA Systems", Genetica, 96, 1995, 89-98. Greenwalt, T. et al., "The Quantification of Fetomaternal Hemorrhage by an Enzyme-Linked Antibody Test with Glutaraldehyde Fixation" Vox Sang. 63, 1992, 268-271.

Fixation", Vox Sang, 63, 1992, 268-271. Grskovic, M. et al., "Validation of a Clinical-Grade Assay to Measure Donor-Derived Cell-Free DNA in Solid Organ Transplant Recipients", The Journal of Molecular Diagnostics, vol. 18, No. 6+Supplemental Appendix S1, Nov. 2016, 890-902.

Guerra, J., "Terminal Contributions for Duplex Oligonucleotide Thermodynamic Properties in the Context of Nearest Neighbor Models", Biopolymers, 95(3), (2010), 2011, 194-201.

Guetta, Esther et al., "Analysis of Fetal Blood Cells in the Maternal Circulation: Challenges, Ongoing Efforts, and Potential Solutions", Stem Cells and Development, 13, 2004, 93-99.

Guichoux, et al., "Current Trends in Microsatellite Genotyping", Molecular Ecology Resources, 11, 2011, 591-911.

Gunderson, K. L. et al., "A genome-wide scalable SNP genotyping assay using microarray technology", Nature Genetics, vol. 37, No. 5, May 2005, 549-554.

Gundry, C. N. et al., "Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplicons", Nucleic Acids Research, vol. 36, No. 10, Apr. 29, 2008, 3401-3408. Hall, M., "Panorama Non-Invasive Prenatal Screening for Microdeletion Syndromes", Apr. 1, 2014 (Apr. 1, 2014), XP055157224, Retrieved from the Internet: URL:http://www.panoramatest.com/sites/default/files/flas/PanoramaMicrodeletionsWhite Paper-2.pdf [retrieved on Dec. 8, 2014].

Han, S-W et al., "Predictive and Prognostic Impact of Epidermal Growth Factor Receptor Mutation in Non-Small-Cell Lung Cancer Patients Treated With Gefitinib", Journal of Clinical Oncology, vol. 23, No. 11, Apr. 10, 2005, 2493-2501.

Handyside, et al., "Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease", Molecular Human Reproduction vol. 10, No. 10 pp. 767-772, 2004.

Hara, Eiji et al., "Subtractive eDNA cloning using oligo(dT)3o-latex and PCR: isolation of eDNA clones specific to undifferentiated human embryonal carcinoma cells", Nucleic Acids Research, 19(25), 1991, 7097-7104.

Hardenbol, P., "Multiplexed Genotyping With Sequence-Tagged Molecular Inversion Probes", Nature Biotechnology, 21 (6), 2003, 673-678.

Hardenbol, Paul et al., "Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a singled tube assay", Genome Research, 15, 2005, 269-275.

Harismendy, O. et al., "Method for Improving Sequence Coverage Uniformity of Targeted Genomic Intervals Amplified by LR-PCR Using Illumina GA Sequencing-By-Synthesis Technology", Bio Techniques, 46(3), 2009, 229-231.

Harper, J. C. et al., "Recent Advances and Future Developments in PGD", Prenatal Diagnosis, 19, 1999, 1193-1199. Harton, G.L. et al., "Preimplantation Genetic Testing for Marfan

Harton, G.L. et al., "Preimplantation Genetic Testing for Marfan Syndrome", Molecular Human Reproduction, 2 (9), 1996, 713-715. Hartwell, L. H. et al., "Chapter 11: The Direct Detection of Genotype Distinguishes Individual Genomes", Genetics: From Genes to Genomes, Second Edition, 2004, 371-414.

Hartwell, L. H. et al., "Chapter 13: Chromosomal Rearrangements and Changes in Chromosome Number Reshape Eukaryotic Genomes", Genetics: From Genes to Genomes, Second Edition, 2004, 441-486. Hattori, M. et al., "The DNA sequence of human chromosome 21", Nature, vol. 405, May 18, 2000, 311-319.

Hawkins, T. et al., "Whole genome amplification—applications and advances", Current Opinion in Biotechnology, 13, 2002, 65-67. Hayden, et al., "Multiplex-Ready PCR: A new method for multiplexed SSR AND SNP genotyping", BMC Genomics 2008, 9(80),

He, Qz et al., "A method for improving the accuracy of non-invasive prenatal screening by cell-free foetal DNA size selection", British Journal of Biomedical science, vol. 75, No. 3, Jul. 2018, 133-138. Hellani, A. et al., "Clinical Application of Multiple Displacement Amplification in Preimplantation Genetic Diagnosis", Reproductive BioMedicine Online, 10 (3), 2005, 376-380.

BioMedicine Online, 10 (3), 2005, 376-380.

Hellani, Ali et al., "Multiple displacement amplification on single cell and possible PGD applications", Molecular Human Reproduction, 10(11), 2004, 847-852.

Hojsgaard, S. et al., "BIFROST—Block recursive models induced

Hojsgaard, S. et al., "BIFROST—Block recursive models induced from relevant knowledge, observations, and statistical techniques", Computational Statistics & Data Analysis, 19(2), 1995, 155-175. Hollas, B. et al., "A stochastic approach to count RN A molecules using DNA sequencing methods", Lecture Notes in Computer Science, vol. 2812, 2003, 55-62.

Holleley, et al., "Multiplex Manager 1.0: a Cross-Platform Computer Program that Plans and Optimizes Multiplex PCR", BioTechniques46:511-517 (Jun. 2009), 511-517.

Hollox, E. et al., "Extensive Normal Copy Number Variation of a  $\beta$ -Defensin Antimicrobial-Gene Cluster", Am. J. Hum. Genet., 73, 2003, 591-600.

Homer, et al., "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays", PLOS Genetics, 4(8), 2008, 9 pgs.

Hoogendoorn, Bastiaan et al., "Genotyping Single Nucleotide Polymorphisms by Primer Extension and High Performance Liquid Chromatography", Hum Genet, 104, 1999, 89-93. Hornak, M. et al., "Aneuploidy Detection in Pigs Using Compara-

Hornak, M. et al., "Aneuploidy Detection in Pigs Using Comparative Genomic Hybridization: From the Oocytes to Blastocysts", PLoS One, vol. 7, No. 1, Jan. 2012, 6 pages.

Hospital, F. et al., "A General Algorithm to Compute Multilocus

Hospital, F. et al., "A General Algorithm to Compute Multilocus Genotype Frequencies Under Various Mating Systems" vol. 12, No. 6, Jan. 1, 1996 (Jan. 1, 1996), pp. 455-462.

## US 11,519,035 B2

Page 12

### (56) References Cited

#### OTHER PUBLICATIONS

Howie, et al., "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing", Nature Genetics, vol. 44, No. 8, Jul. 22, 2012, 955-959.

Hu, Dong Gui et al., "Aneuploidy Detection in Single Cells Using DNA Array-Based Comparative Genomic Hybridization", Molecular Human Reproduction, 10(4), 2004, 283-289.

Hug, H. et al., "Measurement of the number of molecules of a single mRNA species in a complex mRNA preparation", J. Theor. Biol., vol. 221, 2003, 615-624.

Hultin, E. et al., "Competitive enzymatic reaction to control allelespecific extensions", Nucleic Acids Research, vol. 33, No. 5, Mar. 14, 2005, 1-10.

Ido, Yasuo et al., "Hyperglycemia-Induced Apoptosis in Human Umbilical Vein Endothelial Cells: Inhibition by the AMP-Activated Protein Kinase Activation", Diabetes, 51, 2002, 159-167.

Illumina, , "Patent Owner Illumina's Preliminary Response to Petition", Oct. 17, 2018, 75 pgs.

Illumina, , "Petition for Inter Partes Review of U.S. Pat. No. 8,682,592", Jun. 13, 2019, 91 pages.

Illumina,, "Plaintiff/Counterclaim Defendant Illumina, Inc.'s Amended Patent L.R. 3-3 Preliminary Invalidity Contentions for U.S. Pat. No. 8,682,592", Oct. 30, 2018, 22.

Illumina, , "Plaintiff/Counterclaim-Defendant Illumina, Inc.'S Patent L.R. 3-3 Contentions for U.S. Patent Preliminary Invalidity Contentions for U.S. Patent No. 8,682,592", Oct. 9, 2018, 81 pages. Illumina Catalog., "Paired-End Sample Preparation Guide, Illumina Catalog# PE-930-1 001, Part# 1005063 Rev. E", 2011, 1-40.

Illumina, Inc., , "Declaration of David Peters, Ph.D. in Support of Petition for Inter Partes Review of U.S. Pat. No. 8,682,592", Jun. 13, 2019, 136 pages.

Illumina, Inc. V. Natera, Inc., , "Order Re: Claim Construction", Jan. 30, 2019, 16 pgs.

Imielinski, M. et al., "Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing", Cell, vol. 150, Sep. 14, 2012, 1107-1120.

Ishii, et al., "Optimization of Annealing Temperature to Reduce Bias Caused by a Primer Mismatch in Multitemplate PCR", Applied and Environmental Microbiology, Aug. 2001, p. 3753-3755.

Jabara, C. B. et al., "Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID", PNAS, vol. 108, No. 50, Dec. 13, 2011, 20166-20171

Dec. 13, 2011, 20166-20171.

Jahr, S. et al., "DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells", Cancer Research, vol. 61, Feb. 15, 2001, 1659-1665

Jamal-Hanjani, M. et al., "Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer", Annals of Oncology, vol. 27, No. 5, Jan. 28, 2016, 862-867.

Jamal-Hanjani, M. et al., "Tracking Genomic Cancer Evolution for Precision Medicine: The Lung TRACERx Study", PLOS Biology, vol. 12, No. 7, Jul. 2014, 1-7. Jamal-Hanjani, M. et al., "Tracking the Evolution of Non-Small-

Jamal-Hanjani, M. et al., "Tracking the Evolution of Non-Small-Cell Lung Cancer", The New England Journal of Medicine, vol. 376, No. 22, Jun. 1, 2017, 2109-2121.

Jarvie, T., "Next generation sequencing technologies", Drug Discovery Today: Technologies, vol. 2, No. 3, 2005, 255-260.

Jenkins, S. et al., "High-throughput SNP genotyping", Comparative and Functional Genomics, vol. 3, Dec. 5, 2001, 57-66.

Johnson, D.S. et al., "Comprehensive Analysis of Karyotypic Mosaicism Between Trophectoderm and Inner Cell Mass", Molecular Human Reproduction, 16(12), 2010, 944-949.

Johnson D.S, et al., "Preclinical Validation of a Microarray Method for Full Molecular Karyotyping of Blastomeres in a 24-h Protocol", Human Reproduction, 25 (4), 2010, 1066-1075.

Kamat, A. A. et al., "Quantification of total plasma cell-free DNA in ovarian cancer using real-time PCR", Ann N Y Acad Sci., vol. 1075, Sep. 2006, 230-234.

Kaplinski, Lauris et al., "MultiPLX: Automatic Grouping and Evaluation of PCR Primers", Bioinformatics, 21(8), 2005, 1701-1702

Kazakov, V.I. et al., "Extracellular DNA in the Blood of Pregnant Women", Tsitologia, vol. 37, No. 3, 1995, 1-8.

Kijak, G. et al., "Discrepant Results in the Interpretation of HIV-1 Drug-Resistance Genotypic Data Among Widely Used Algorithms", HIV Medicine, 4, 2003, 72-78.

Kim, H. et al., "Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution", Genome Research, vol. 25, No. 3, Feb. 3, 2015, 316-327.

Kinde, I. et al., "Detection and quantification of rare mutations with massively parallel sequencing", PNAS, vol. 108, No. 23, Jun. 7, 2011, 9530-9535.

Kinnings, S. L. et al., "Factors affecting levels of circulating cell-free fetal DNA in maternal plasma and their implications for noninvasive prenatal testing", Prenatal Diagnosis, vol. 35, 2015, 816-822.

Kirkizlar, E. et al., "Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology", Translational Oncology, vol. 8, No. 5, Oct. 2015, pp. 407-416.

Kivioja, T. et al., "Counting absolute numbers of molecules using unique molecular identifiers", Nature Methods, Advance Online Publication, Nov. 20, 2011, 1-5.

Konfortov, Bernard A. et al., "An Efficient Method for Multi-Locus Molecular Haplotyping", Nucleic Acids Research, 35(1), E6, 2007, 8 pgs.

Krjutskov, K. et al., "Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays", Nucleic Acids Research, vol. 36, No. 12, May 23, 2008, 7 pages.

Kuliev, Anver et al., "Thirteen Years' Experience on Preimplantation Diagnosis: Report of the Fifth International Symposium on Preimplantation Genetics", Reproductive BioMedicine Online, 8, 2, 2004, 229-235.

Kunishima, S. et al., "First description of somatic mosaicism in MYH9 disorders", British Journal of Haematology, vol. 128, 2005, 360-365.

Kwok, P. Y., "High-throughput genotyping assay approaches", Pharmacogenomics, vol. 1, No. 1, 2000, 1-5.

Lambert-Messerlian, G. et al., "Adjustment of Serum Markers in First Trimester Screening", Journal of Medical Screening, 16 (2), 2009, 102-103.

Lander, E. S. et al., "Initial sequencing and analysis of the human genome", Nature, vol. 409, Feb. 15, 2001, 860-921. Lathi, Ruth B. et al., "Informatics Enhanced SNP Microarray

Lathi, Ruth B. et al., "Informatics Enhanced SNP Microarray Analysis of 30 Miscarriage Samples Compared to Routine Cytogenetics", PLoS One, 7(3), 2012, 5 pgs.

Leary, R. J. et al., "Development of Personalized Tumor Biomarkers

Leary, R. J. et al., "Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing", Science Translational Medicine, vol. 2, No. 20, Feb. 24, 2010, 1-8.

Leary, Rebecca J et al., "Detection of Chromosomal Alterations in the Circulation of Cancer Patients with Whole-Genome Sequencing", Science Translational Medicine, 4, 162, 2012, 12.

Levsky, J. M. et al., "Fluorescence in situ hybridization: past, present and future", Journal of Cell Science, vol. 116, No. 14, 2003, 2833-2838.

Li, B. , "Highly Multiplexed Amplicon Preparation for Targeted Re-Sequencing of Sample Limited Specimens Using the Ion AmpliSeq Technology and Semiconductor Sequencing", Proceedings of the Annual Meeting of the American Society of Human Genetics [retrieved on Oct. 30, 2012]. Retrieved from the Internet: <URL: http://www.ashg.org/2012meeting/abstracts/fulltext/f120121811. htm>, 2012, 1 pg.

Li, R. et al., "SNP detection for massively parallel whole-genome resequencing", Genome Research, vol. 19, 2009, 1124-1132. Li, Y. et al., "Non-Invasive Prenatal Diagnosis Using Cell-Free

Li, Y. et al., "Non-Invasive Prenatal Diagnosis Using Cell-Free Fetal DNA in Maternal Plasma from PGD Pregnancies", Reproductive BioMedicine Online, 19 (5), 2009, 714-720.

## US 11,519,035 B2

Page 13

#### (56) References Cited

#### OTHER PUBLICATIONS

Li, Ying et al., "Size Separation of Circulatory DNA in Maternal Plasma Permits Ready Detection of Fetal DNA Polymorphisms", Clinical Chemistry, 50, 6, 2004, 1002-1011.

Liao, Gary J.W. et al., "Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles", Clinical Chemistry, 57 (1), 2011, 92-101. Liao, J. et al., "An Alternative Linker-Mediated Polymerase Chain

Liao, J. et al., "An Alternative Linker-Mediated Polymerase Chain Reaction Method Using a Dideoxynucleotide to Reduce Amplification Background", Analytical Biochemistry 253, 137-139 (1997). Liew, Michael et al., "Genotyping of Single-Nucleotide Polymorphisms", Clinical Chemistry, 50(7), 2004, 1156-1164. Life Technologies, , "Ion AmpliSeq<sup>TM</sup> Designer provides full flex-

ibility to sequence genes of your choice?, Retrieved from the Internet<URL: http://tools.lifetechnologies.com/content/sfs/brochures/IonAmpliSeq\_CustomPanels\_AppNote\_CO1, 2012, 4 pages. Lindberg, J. et al., "Exome Sequencing of Prostate Cancer Supports

Lindberg, J. et al., "Exome Sequencing of Prostate Cancer Supports the Hypothesis of Independent Tumour Origins", European Urology, vol. 63, 2013, 347-353.
Lindroos, Katatina et al., "Genotyping SNPs by Minisequencing Primer Extension Using Oligonucleotide Microarrays", Methods in Molecular Biology, 212, Single Nucleotide Polymorphisms: Methods and Protocols, P-K Kwok (ed.), Humana Press, Inc., Totowa,

NJ, 2003, 149-165. Lo, et al., "Digital PCR for the Molecular Detection of Fetal Chromosomal Aneuploidy", PNAS, vol. 104, No. 32, Aug. 7, 2007, 13116-13121.

Lo, et al., "Fetal Nucleic Acids in Maternal Blood: the Promises", Clin. Chem. Lab. Med., 50(6), 2012, 995-998.

Lo, et al., "Free Fetal DNA in Maternal Circulation", JAMA, 292(23), (Letters to the Editor), 2004, 2835-2836.

Lo, , "Non-Invasive Prenatal Diagnosis by Massively parallel Sequencing of Maternal Plasma DNA", Open Biol 2: 120086, 2012, 1-5

Lo, et al., "Prenatal Sex Determination by DNA Amplification from Maternal Peripheral Blood", The Lancet, 2, 8676, 1989, 1363-1365. Lo, et al., "Rapid Clearance of Fetal DNA from Maternal Plasma", Am. J. Hum. Genet., 64, 1999, 218-224.

Lo, et al., "Strategies for the Detection of Autosomal Fetal DNA Sequence from Maternal Peripheral Blood", Annals New York Academy of Sciences,731, 1994, 204-213.

Lo, et al., "Two-way cell traffic between mother and fetus: biologic and clinical implications", Blood, 88(11), Dec. 1, 1996, 4390-4395. Lo, Y., "Noninvasive prenatal detection of fetal chromosomal aneuploidies by maternal plasma nucleic acid analysis: a review of the current state of the art", BJOG An International Journal of Obstetrics and Gynaecology, vol. 116, 2009, 152-157.

Lo, Y.M. Dennis, "Fetal Nucleic Acids in Maternal Plasma: Toward the Development of Noninvasive Prenatal Diagnosis of Fetal Chromosomal Aneuploidies", Ann. N.Y. Acad. Sci., 1137, 2008, 140-143.

Lo, Y.M. Dennis et al., "Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus", Science Translational Medicine, 2 (61), 2010, 13.

Lo, Y.M. Dennis et al., "Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection", Nature Medicine, 13 (2), 2007, 218-223.

Lo, Y.M. Dennis et al., "Presence of Fetal DNA in Maternal Plasma and Serum", The Lancet, 350, 1997, 485-487.

Lo, Y.M. Dennis et al., "Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis", Am. J. Hum. Genet., 62, 1998, 768-775.

Lo, Y-M D., "Non-invasive prenatal diagnosis using fetal cells in maternal blood", J. Clin. Pathol., vol. 47, 1994, 1060-1065.

Lo, Y-M.D et al., "Detection of Single-Copy Fetal DNA Sequence from Maternal Blood", The Lancet, 335, 1990, 1463-1464.

Lo, Y-M.D et al., "Prenatal Determination of Fetal Rhesus D Status by DNA Amplification of Peripheral Blood of Rhesus-Negative Mothers", Annals New York Academy of Sciences, 731, 1994, 229-236. Lo, Y-M.D. et al., "Detection of Fetal RhD Sequence from Peripheral Blood of Sensitized RhD-Negative Pregnant Women", British Journal of Haematology, 87, 1994, 658-660.

Lo, Y-M.D. et al., "Prenatal Determination of Fetal RhD Status by Analysis of Peripheral Blood of Rhesus Negative Mothers", The Lancet, 341, 1993, 1147-1148.

Lu, I. et al., "Establishment of a system based on universal multiplex-PCR for screening genetically modified crops", Anal. Bioanal. Chem. vol. 396, Oct. 24, 2009, 2055-2064.

Lui, Y. Y. et al., "Predominant Hematopoietic Origin of Cell-Free DNA in Plasma and Serum after Sex-Mismatched Bone Marrow Transplantation", Clinical Chemistry, vol. 48, vol. 3, 2002, 421-427. Lun, Fiona M. et al., "Noninvasive Prenatal Diagnosis of Monogenic Diseases by Digital Size Selection and Relative Mutation Dosage on DNA in Maternal Plasma", PNAS, 105(50), 2008, 19920-19925.

Ma, Xiaotu et al., "Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia", Nature Communications, vol. 6, Mar. 19, 2015, 1-12.

Communications, vol. 6, Mar. 19, 2015, 1-12. Magbanua, M. J. et al., "Abstract PD2-01: Personalized serial circulating tumor DNA (ctDNA) analysis in high-risk early stage breast cancer patients to monitor and predict response to neoadjuvant therapy and outcome in the I-SPY 2 Trial", Cancer Research, vol. 79, No. 4 Supplement, Feb. 15, 2019.

Mamon, H. et al., "Letters to the Editor: Preferential Amplification of Apoptotic DNA from Plasma: Potential for Enhancing Detection of Minor DNA Alterations in Circulating DNA", Clinical Chemistry, vol. 54, No. 9, 2008, 1582-1584.

Maniatis, T. et al., "In: Molecular Cloning: A Laboratory Manual",

Maniatis, T. et al., "In: Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory, New York, Thirteenth Printing, 1986, 458-459.

Mansfield, Elaine S , "Diagnosis of Down Syndrome and Other Aneuploidies Using Quantitative Polymerase Chain Reaction and Small Tandem Repeat Polymorphisms", Human Molecular Genetics, 2, 1, 1993, 43-50.

Mardis, E. R., "The impact of next-generation sequencing technology on genetics", Trends in Genetics, vol. 24, No. 3, Feb. 11, 2008, 133-141.

Margulies, M. et al., "Genome sequencing in microfabricated highdensity picolitre reactors", Nature, vol. 437, Sep. 15, 2005, 376-380. Margulies, M. et al., "Genome sequencing in microfabricated highdensity picolitre reactors plus Supplemental Methods", Nature, vol. 437, Sep. 15, 2005, 40 pgs. Markoulatos, P. et al., "Multiplex Polymerase Chain Reaction: A

Markoulatos, P. et al., "Multiplex Polymerase Chain Reaction: A Practical Approach", Journal of Clinical Laboratory Analysis, vol. 16, 2002, 47-51.

May, Robert M., "How Many Species Are There on Earth?", Science, 241, Sep. 16, 1988, 1441-1449.

Science, 241, Sep. 16, 1988, 1441-1449.
McBride, D. et al., "Use of Cancer-Specific Genomic Rearrangements to Quantify Disease Burden in Plasma from Patients with Solid Tumors", Genes, Chromosomes & Cancer, vol. 49, Aug. 19, 2010, 1062-1069.

McCloskey, M. L. et al., "Encoding PCR Products with Batch-stamps and Barcodes", Biochem Genet., vol. 45, Oct. 23, 2007, 761-767.

McCray, Alexa T. et al., "Aggregating UMLS Semantic Types for Reducing Conceptual Complexity", MEDINFO 2001: Proceedings of the 10th World Congress on Medical Informatics (Studies in Health Technology and Informatics, 84, V. Patel et al. (eds.), IOS Press Amsterdam, 2001, 216-220.

McDonald, B. R. et al., "Abstract P4-01-21: Multiplexed targeted digital sequencing of circulating tumor DNA to detect minimal residual disease in early and locally advanced breast cancer", Cancer Research, vol. 79, No. 4 Supplement, Feb. 15, 2019.

Mennuti, M. et al., "Is It Time to Sound an Alarm About False-Positive Cell-Free DNA Testing for Fetal Aneuploidy?", American Journal of Obstetrics, 2013, 5 pgs.

Merriam-Webster, "Medical Definition of Stimulant", http://www.merriam-webster.com/medical/stimulant, Mar. 14, 2016, 7 pages. Merriam-Webster, , "Universal Definition", "Merriam-Webster.com (http://www.merriam-webster.com/dictionary/universal, downloaded Jul. 23, 2014)", 3 pages.

Mersy, et al., "Noninvasive Detection of Fetal Trisomy 21: Systematic Review and Report of Quality and Outcomes of Diagnostic

Case: 24-1324 Document: 42-1 Page: 49 Filed: 03/18/2024

## US 11,519,035 B2

Page 14

#### (56)References Cited

#### OTHER PUBLICATIONS

Accuracy Studies Performed Between 1997 and 2012", Human Reproduction Update, 19(4), 2013, 318-329.

Mertes, F. et al., "Targeted enrichment of genomic DNA regions for next-generation sequencing", Briefings in Functional Genomics, vol. 10, No. 6, Nov. 26, 2011, 374-386

Miller, Robert, "Hyperglycemia-Induced Changes in Hepatic Membrane Fatty Acid Composition Correlate with Increased Caspase-3 Activities and Reduced Chick Embryo Viability", Comparative

Biochemistry and Physiology, Part B, 141, 2005, 323-330. Miller, Robert R., "Homocysteine-Induced Changes in Brain Membrane Composition Correlate with Increased Brain Caspase-3 Activities and Reduced Chick Embryo Viability", Comparative Biochemistry and Physiology Part B, 136, 2003, 521-532.

Miner, B. E. et al., "Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR", Nucleic Acids Research, vol. 32, No. 17, Sep. 30, 2004, 1-4. Minkoff, E. et al., "Stem Cells, Cell Division, and Cancer", Biology

Today Third Edition, Chapter 12, 2004, 10 pages.

Morand, et al., "Hesperidin contributes to the vascular protective effects of orange juice: a randomized crossover study in healthy volunteers", Am J Clin Nutr. Jan. 2011;93(1):73-80. Epub Nov. 10, 2010.

Munne, S. et al., "Chromosome Abnormalities in Human Embryos". Textbook of Assisted Reproductive Techniques, 2004, pp. 355-377. Munne, S. et al., "Chromosome abnormalities in human embryos", European Society of Human Reproduction and Embryology: Human Reproduction Update, vol. 4, No. 6, 1998, 842-855.

Munne, S. et al., "Improved implantation after preimplantation genetic diagnosis of aneuploidy", Reproductive BioMedicine Online, vol. 7., No. 1., May 15, 2003, 91-97. Murtaza, M. et al., "Non-Invasive Analysis of Acquired Resistance

to Cancer Therapy by Sequencing of Plasma DNA", Nature (doi:10. 1038/nature12065), 2013, 6 pgs.

Muse, Spencer V., "Examining rates and patterns of nucleotide substitution in plants", Plant Molecular Biology 42: 25-43, 2000. Myers, Chad L. et al., "Accurate Detection of Aneuploidies in Array CGH and Gene Expression Microarray Data", Bioinformatics, 20(18), 2004, 3533-3543.

Nannya, Yasuhito et al., "A Robust Algorithm for Copy Number Detection Using High-density Oligonucleotide Single Nucleotide Polymorphism Genotyping Arrays", Cancer Res., 65, 14, 2005,

Narayan, A. et al., "Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing", Cancer Research, vol. 72, No. 14, Jul. 15, 2012, 3492-3498

Natera, Inc., "Declaration of Sandra L. Haberny", May 16, 2019, 3 pages

Natera, Inc., , "Defendant Natera, Inc.'s Invalidity Contentions Under Patent L.R. 3-3; Document Production Accompanying Invalidity Contentions Under Patent L.R. 3-4", Aug. 20, 2018, 17 pages. Natera, Inc., , "Exhibit 8 EHRICH Invalidity Chart", Aug. 20, 2018, 16 pages.

Natera, Inc., "Exhibits A-H to Haberny Declaration", May 16, 2019, 192 pages.

Natera, Inc., "Motion to Dismiss", May 16, 2019, 2 pages. Natera, Inc., "Natera Inc.'s First Amended Answer, Affirmative

Defenses and Counterclaims", Aug. 16, 2018, 28 pages

Natera, Inc., "Natera, Inc.'s Supplemental Objections and Response to Plaintiff Illumina, Inc.'s Interrogatory No. 8", Mar. 20, 2019, 29 pages.

Natera, Inc., , "Opening Brief in Support of Motion to Dismiss", May 16, 2019, 26 pages.

Natera, Inc., , "Petitioner Reply Per Board Order of Nov. 2, 2018 (Paper No. 10)", Nov. 9, 2018, 8 pgs.

NCBI, , "dbSNP record for rs1294331", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs 1294331 >, 2019, 2

NCBI, , "dbSNP record for rs1872575", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs1872575, 2019, 2 pgs.</p> NCBI, , "dbSNP record for rs2362450", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs2362450>, 2019, 1 pg. NCBI, , "dbSNP record for rs2384571", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs2384571>, 2019, 2 pgs. NCBI, , "dbSNP record for rs2498982", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs2498982>, 2019, 3 pgs. NCBI, , "dbSNP record for rs3731877", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.govisnp/?term=rs3731877>, 2019, 2 pgs. Newman, A. M. et al., "Integrated digital error suppression for improved detection of circulating tumor DNA", Nature Biotechnology, vol. 34, No. 5, May 2016, 547-555. Ng, S. B. et al., "Individualised multiplexed circulating tumour

DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery", Scientific Reports, vol. 7, No. 40737, Jan. 19, 2017, 11 pages.

Nguyen-Dumont, T., "A high-plex PCR approach for massively parallel sequencing", BioTechniques, vol. 55, No. 2, Aug. 2013, 69-74.

Nicolaides, K. et al., "Noninvasive Prenatal Testing for Fetal Trisomies in a Routinely Screened First-Trimester Population", American Journal of Obstetrics (article in press), 207, 2012, 1.e1-

Nicolaides, K.H et al., "Validation of Targeted Sequencing of Single-Nucleotide Polymorphisms for Non-Invasive Prenatal Detection of Aneuploidy of Chromosomes 13, 18, 21, X, and Y", Prenatal Diagnosis, 33, 2013, 575-579.

Nicolaides, Kypros H. et al., "Prenatal Detection of Fetal Triploidy from Cell-Free DNA Testing in Maternal Blood", Fetal Diagnosis and Therapy, 2013, 1-6.

Nygren, et al., "Quantification of Fetal DNA by Use of Methylation-Based DNA Discrimination", Clinical Chemistry 56:10 1627-1635

Ogino, S. et al., "Bayesian Analysis and Risk Assessment in Genetic Counseling and Testing", Journal of Molecular Diagnostics, 6 (1), 2004, 9 pgs.

Ohsawa, M. et al., "Prenatal Diagnosis of Two Pedigrees of Fukuyama Type Congenital Muscular Dystrophy by Polymorphism Analysis", The Health and Welfare Ministry, 1994, 5 pgs.

O'Malley, R. et al., "An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome", Nat. Protoc., 2, 2007, 2910-2917.

Orozco A.F., et al., "Placental Release of Distinct DNA-Associated Micro-Particles into Maternal Circulation: Reflective of Gestation

Time and Preeclampsia", Placenta, 30, 2009, 891-897. Ozawa, Makiko et al., "Two Families with Fukuyama Congenital Muscular Dystrophy that Underwent in Utero Diagnosis Based on Polymorphism Analysis", Clinical Muscular Dystrophy: Research in Immunology and Genetic Counseling—FY 1994 Research Report,

(including text in Japanese), 1994, 8.
Paez, Guillermo J. et al., "Genome coverage and sequence fidelity of \$49 polymerase-based multiple strand displacement whole genome amplification", Nucleic Acids Research, 32(9), 2004, 1-11

Page, S. L. et al., "Chromosome Choreography: The Meiotic Ballet", Science, 301, 2003, 785-789.

Palomaki, G. E. et al., "DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study", Genetics in Medicine, vol. 13, No. 1, Nov. 2011, 913-920.

Palomaki, Glenn et al., "DNA Sequencing of Maternal Plasma Reliably Identifies Trisomy 18 and Trisomy 13 as Well as Down Syndrome: an International Collaborative Study", Genetics in Medicine, 2012, 10.

Palomaki, Glenn E. et al., "DNA Sequencing of Maternal Plasma to Detect Down Syndrome: An International Clinical Validation Study", Genetics in Medicine (pre-print version), 13, 2011, 8 pgs

Papadopoulou, E. et al., "Cell-Free DNA and RNA in Plasma as a New Molecular Marker for Prostate Cancer", Oncology Research, vol. 14, 2004, 439-445.

Papageorgiou, Elisavet A. et al., "Fetal-Specific DNA Methylation Ratio Permits Noninvasive Prenatal Diagnosis of Trisomy 21", Nature Medicine (advance online publication), 17, 2011, 5 pgs.

### US 11,519,035 B2

Page 15

#### (56) References Cited

#### OTHER PUBLICATIONS

Pastinen, T. et al., "Minisequencing: A Specific Tool for DNA Analysis and Diagnostics on Oligonucleotide Arrays", Genome Research, vol. 7, 1997, 606-614.

Pathak, A. et al., "Circulating Cell-Free DNA in Plasma/Serum of Lung Cancer Patients as a Potential Screening and Prognostic Tool", Clinical Chemistry, 52, 2006, 1833-1842.

PCT/US2006/045281, , "International Preliminary Report on Patentability", dated May 27, 2008, 1 pg.

PCT/US2006/045281, , "International Search Report and Written Opinion", dated Sep. 28, 2007, 7 pgs.

PCT/US2008/003547, , "International Search Report", dated Apr.  $15,\,2009,\,5$  pgs.

 $PCT/US2009/034506, , "International Search Report", dated Jul.\ 8, 2009,\ 2\ pgs.$ 

PCT/US2009/045335, , "International Search Report", dated Jul. 27, 2009, 1 pg.

PCT/US2009/052730, , "International Search Report", dated Sep. 28, 2009, 1 pg.

PCT/US2010/050824, "International Search Report", dated Nov. 15, 2010, 2 pgs.

PCT/US2011/037018, , "International Search Report", dated Sep.

27, 2011, 2 pgs. PCT/US2011/061506, , "International Search Report", dated Mar.

16, 2012, 1 pgs.

PCT/US2011/066938, , "International Search Report", dated Jun. 20, 2012, 1 pg.

PCT/US2012066339, , "International Search Report", dated Mar. 5, 2013, 1 pg.

PCT/US2013/028378, , "International Search Report and Written Opinion", dated May 28, 2013, 11 pgs.

PCT/US2013/57924, "International Search Report and Written Opinion", dated Feb. 18, 2014, 8 pgs.

PCT/US2014/051926, , "International Search Report and Written Opinion", dated Dec. 9, 2014, 3 pgs.

Pearson, K., "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", Philosophical Magazine Series 5, vol. 50, Issue 302, 1900, 157-175.

Pena, Sergio D.J et al., "Paternity Testing in the DNA Era", Trends in Genetics, 10, 6, 1994, 204-209.

Pergament, E. et al., "Single-Nucleotide Polymorphism-Based Noninvasive Prenatal Screening in a High-Risk and Low-Risk Cohort", Obstetrics & Gynecology, vol. 124, No. 2, Part 1, Aug. 2014, 210-218+Appendices.

210-218+Appendices.

Perkel, Jeffrey M., "Overcoming the Challenges of Multiplex PCR" Biocompare Editorial Article 2012, 1-5.

PCR", Biocompare Editorial Article, 2012, 1-5.
Perry, George H. et al., "The Fine-Scale and Complex Architecture of Human Copy-Number Variation", The American Journal of Human Genetics, 82, 2008, 685-695.

Human Genetics, 82, 2008, 685-695.
Pertl, B. et al., "Detection of Male and Female Fetal DNA in Maternal Plasma by Multiplex Fluorescent Polymerase Chain Reaction Amplification of Short Tandem Repeats", Hum. Genet., 106, 2000, 45-49.

Peters, D., "List of Materials Considered by David Peters, Ph.D.", Jun. 13, 2019, 2 pages.

Peters, David P. et al., "Noninvasive Prenatal Diagnosis of a Fetal Microdeletion Syndrome", New England Journal of Medicine, 365(19), 2011, 1847-1848.

Pfaffl, Michael W., "Relative Expression Software Tool (REST©) for Group-Wise Comparison and Statistical Analysis of Relative Expression Results in real-Time PCR", Nucleic Acids Research, 30(9), 2002, 10 pgs.

Phillips, C. et al., "Resolving Relationship Tests that Show Ambiguous STR Results Using Autosomal SNPs as Supplementary Markers", Forensic Science International: Genetics 2, 2008, 198-204. Pirker, C. et al., "Whole Genome Amplification for CGH Analysis:

Linker-Adapter PCR as the Method of Choice for Difficult and Limited Samples", Cytometry Part A, vol. 61A, 2004, 26-34. Podder, Mohua et al., "Robust SN P genotyping by multiplex PCR and arrayed primer", BMC Medical Genomics,1(5), 2008, 1-15. Poirier, K. et al., "Maternal mosaicism for mutations in the ARX gene in a family with X linked mental retardation", Human Genetics, vol. 118, Aug. 3, 2005, 45-48. Poon, L. L. et al., "Differential DNA Methylation between Fetus and

Poon, L. L. et al., "Differential DNA Methylation between Fetus and Mother as a Strategy for Detecting Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 48, No. 1, 2002, 35-41.

Popova, T. et al., "Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays", Genome Biology, vol. 10, R128, Nov. 11, 2009, 1-14. Porreca, Gregory J et al., "Multiplex Amplification of Large Sets of Human Exons", Nature Methods, 4, (advance online publication), 2007. 6.

Price, T.S. et al., ""SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data",", Nucleic Acids Research, vol. 33, No. 11, Jun. 16, 2005, 3455-3464.

Primdahl, H. et al., "Allelic Imbalances in Human Bladder Cancer: Genome-Wide Detection With High-Density Single-Nucleotide Polymorphism Arrays", Journal of the National Cancer Institute, vol. 94, No. 3, Feb. 6, 2002, 216-223.

Quinn, G. P. et al., "Experimental Design and Data Analysis for Biologists", Graphical Exploration of Data, 2002, 64-67.

Rabinowitz, et al., "Accurate Prediction of HIV-1 Drug Response from the Reverse Transcriptase and Protease Amino Acid Sequences Using Sparse Models Created by Convex Optimization", Bioinformatics, 22, 5, 2006, 541-549.

Rabinowitz, Matthew et al., "Origins and rates of aneuploidy inhuman blastomeres", Fertility and Sterility, vol. 97, No. 2, Feb. 2012, 395-401.

Rabinowitz, Matthew. et al., "Non-Invasive Prenatal Aneuploidy Testing of Chromosomes 13, 18, 21, X, and Y Using Targeted Sequencing of Polymorphic Loci", The American Society of Human Genetics, meeting poster, 2012, 1 pg.

Rachlin, J. et al., "Computational tradeoffs in multiplex PCR assay design for SNP genotyping", BMC Genomics, vol. 6, No. 102, Jul. 25, 2005, 11 pages.

Ragoussis, J., "Genotyping Technologies for Genetic Research", Annual Review of Genomics and Human Genetics, vol. 10 (1), Sep. 1, 2009, 117-133.

Rahmann, Sven et al., "Mean and variance of the Gibbs free energy of oligonucleotides in the nearest neighbor model under varying conditions", Bioinformatics, 20(17), 2004, 2928-2933.

Rava, Richard P. et al., "Circulating Fetal Cell-Free DNA Fraction Differ in Autosomal Aneuploidies and Monosomy X", Clinical Chemistry, 60(1), (papers in press), 2013, 8 pgs.
Rechitsky, Svetlana et al., "Preimplantation Genetic Diagnosis with

Rechitsky, Svetlana et al., "Preimplantation Genetic Diagnosis with HLA Matching", Reproductive Bio Medicine Online, 9, 2, 2004, 210-221.

Reinert, T. et al., "Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery", Gut, vol. 65, 2016, 625-634.

Renwick, P. et al., "Proof of Principle and First Cases Using Preimplantation Genetic Haplotyping—A Paradigm Shift for Embryo Diagnosis", Reproductive BioMedicine Online, 13 (1), 2006, 110-119.

Ricciotti, Hope, "Eating by Trimester", [Online]. Retrieved from Internet<a href="http://www.youandyourfamily.com/article.php?story=Eating+by+Trimester">http://www.youandyourfamily.com/article.php?story=Eating+by+Trimester</a>, 2014, 3.

Riley, D. E., "DNA Testing: An Introduction for Non-Scientists An Illustrated Explanation", Scientific Testimony: An Online Journal, http://www.scientific.org/tutorials/articles/riley/riley.html, Apr. 6, 2005, 22 pages.

Riva, F., "Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer", Clinical Chemistry, vol. 63, No. 3, 2017, 691-699.

Rogaeva, E. et al., "The Solved and Unsolved Mysteries of the Genetics of Early-Onset Alzheimer's Disease", NeuroMolecular Medicine, vol. 2, 2002, 1-10.

Roper, Stephen M. et al., "Forensic Aspects of DNA-Based Human Identity Testing", Journal of Forensic Nursing, 4, 2008, 150-156.

### US 11,519,035 B2

Page 16

#### (56) References Cited

#### OTHER PUBLICATIONS

Roux, K., "Optimization and Troubleshooting in PCR", PCR Methods Appl. 4, 1995, 185-194.

Rozen, Steve et al., "Primer3 on the WWW for General Users and for Biologis Programmers", Methods in Molecular Biology, 132: Bioinformatics Methods and Protocols, 1999, 365-386.

Russell, L. M., "X Chromosome Loss and Ageing", Cytogenetic and Genome Res., 116, 2007, 181-185.

Ryan, A. et al., "Informatics-Based, Highly Accurate, Noninvasive Prenatal Paternity Testing", Genetics in Medicine (advance online publication), 2012, 5 pgs.

Rychlik, et al., "Optimization of the annealing temperature for DNA amplification in vitro", Nucleic Acids Research, 18(21), 1990, 6409-6412.

Sahota, A., "Evaluation of Seven PCR-Based Assays for the Analysis of Microchimerism", Clinical Biochemistry, vol. 31, No. 8., 1998, 641-645.

Saker, A. et al., "Genetic characterisation of circulating fetal cells allows non-invasive prenatal diagnosis of cystic fibrosis", Prenatal Diagnosis, vol. 26, Jul. 11, 2006, 906-916.

Samango-Sprouse, C. et al., "SNP-Based Non-Invasive Prenatal Testing Detects Sex Chromosome Aneuploidies with High Accuracy", Prenatal Diagnosis, 33, 2013, 1-7.
Sanchez, C. et al., "New insights into structural features and optimal

Sanchez, C. et al., "New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis", Nature Partner Journals, vol. 3, No. 31, Nov. 23, 2018, 12 pgs.

Sander, Chris, "Genetic Medicine and the Future of Health Care", Science, 287(5460), 2000, 1977-1978.

Santalucia, J. et al., "The Thermodynamics of DNA Structural Motifs", Annu. Rev. Biophys. Biomol. Struct., 33, 2004, 415-440. Santalucia, John J.R et al., "Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability", Biochemistry, 35, 1996, 3555-3562.

Sasabe, Yutaka, "Genetic Diagnosis of Gametes and Embryos Resulting from Art", Japanese Journal of Fertility and Sterility, vol. 46, No. 1, 2001, 43-46.

Schmitt, M. W. et al., "Detection of ultra-rare mutations by next-generation sequencing", PNAS, vol. 109, No. 36, Sep. 4, 2012, 14508-14513.

Schoumans, J et al., "Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH)", JMed Genet, 42, 2005, 699-705.

Schwarzenbach, H. et al., "Cell-free nucleic acids as biomarkers in cancer patients", Nature Reviews: Cancer, vol. 11, Jun. 2011, 426-437.

Sebat, Jonathan et al., "Strong Association of De Novo Copy Number Mutations with Autism", Science, 316, 2007, 445-449. Sehnert, A. et al., "Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood", Clinical Chemistry (papers in press), 57 (7), 2011, 8 pgs.

Sermon, Karen et al., "Preimplantation genetic diagnosis", The Lancet, Lancet Limited. 363(9421), 2000, 1633-1641.

Servin, B et al., "MOM: A Program to Compute Fully Informative Genotype Frequencies in Complex Breeding Schemes", Journal of Heredity, vol. 93, No. 3, Jan. 1, 2002 (Jan. 1, 2002), pp. 227-228. Sham, P. et al., "DNA Pooling: a Tool for Large-Scale Association Studies", Nature Reviews Genetics, vol. 3, Nov 2002, 862-871.

Shaw-Smith, et al., "Microarray Based Comparative Genomic Hybridisation (array-CGH) Detects Submicroscopic Chromosomal Deletions and Duplications in Patients with Learning Disability/Mental Retardation and Dysmorphic Features", J. Med. Genet., 41, 2004, 241-248.

Shen, et al., "High-quality DNA sequence capture of 524 disease candidate genes", High-quality DNA sequence capture of 524 disease candidate genes, Proceedings of the National Academy of Sciences, vol. 108, No. 16, Apr. 5, 2011 (Apr. 5, 2011), pp. 6549-6554.

Shen, R. et al., "High-throughput SNP genotyping on universal bead arrays", Mutation Research, vol. 573, Feb. 11, 2005, 70-82.

Shen, Zhiyong, "MPprimer: a program for reliable multiplex PCR primer design", BMC Bioinformatics 2010, 11:143, 1-7.

Sherlock, J et al., "Assessment of Diagnostic Quantitative Fluorescent Multiplex Polymerase Chain Reaction Assays Performed on Single Cells", Annals of Human Genetics, 62, Jan. 1998, 9-23.

Shi, H. et al., "Melanoma whole-exome sequencing identifies V600E B-RAF amplification-mediated acquired B-RAF inhibitor resistance", Nature Communications, vol. 3, No. 724, Mar. 6, 2012, 8 pages.

Shiroguchi, K. et al., "Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes", PNAS, vol. 109, No. 4, Jan. 4, 2012, 1347-1352

Sigdel, T. et al., "Plasma Donor-Derived Cell-Free DNA Quantification by massively multiplex PCR Distinguishes Kidney Transplant Acute Rejection", Transplantation, vol. 102, No. 7S, Jul. 2018, S178-S179.

Sigdel, T. K. et al., "Optimizing Detection of Kidney Transplant Injury by Assessment of Donor-Derived Cell-Free DNA via Massively Multiplex PCR", Journal of Clinical Medicine, vol. 8, No. 19, Dec. 23, 2018, 17 pages.

Dec. 23, 2018, 17 pages. Simpson, J. et al., "Fetal Cells in Maternal Blood: Overview and Historical Perspective", Annals New York Academy of Sciences, 731, 1994, 1-8.

Sint, Daniela et al., "Advances in Multiplex PCR: Balancing Primer Efficiencies and Improving Detection Success", Methods in Ecology and Evolution, Mar. 2012, 898-905.

Slater, Howard et al., "High-Resolution Identification of Chromosomal Abnormalities Using Oligonucleotide Arrays Containing 116,204 SNPs", Am. J. Hum. Genet., 77, 5, 2005, 709-726.

Snijders, Antoine et al., "Assembly of Microarrays for Genome-Wide Measurement of DNA Copy Number", Nature Genetic, 29, 2001, 263-264.

Snyder, T. M. et al., "Universal noninvasive detection of solid organ transplant rejection", PNAS, vol. 108, No. 15, Apr. 12, 2011, 6229-6234.

Sparks, A. et al., "Non-Invasive Prenatal Detection and Selective Analysis of Cell-Free DNA Obtained from Maternal Blood: Evaluation for Trisomy 21 and Trisomy 18", American Journal of Obstetrics & Gynecology 206, 2012, 319,e1-319.e9.

Obstetrics & Gynecology 206, 2012, 319.e1-319.e9.
Sparks, Andrew B. et al., "Selective Analysis of Cell-Free DNA in Maternal Blood for Evaluation of Fetal Trisomy", Prenatal Diagnosis, 32, 2012, 1-7.

Spertini, D. et al., "Screening of Transgenic Plants by Amplification of Unknown Genomic DNA Flanking T-DNA", BioTechniques, vol. 27, Aug. 1999, 308-314.

Spiro, Alexander et al., "A Bead-Based Method for Multiplexed Identification and Quantitation of DNA Sequences Using Flow Cytometry", Applied and Environmental Microbiology, 66, 10, 2000, 4258-4265.

Spits, C et al., "Optimization and Evaluation of Single-Cell Whole Genome Multiple Displacement Amplification", Human Mutation, 27(5), 496-503, 2006.

Srinivasan, et al., "Noninvasive Detection of Fetal Subchromosome Abnormalities via Deep Sequencing of Maternal Plasma", The American Journal of Human Genetics 92, 167-176, Feb. 7, 2013. Stephens, Mathews. et al., "A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data", Am. J. Hum. Genet.,73, 2003, 1162-1169.

Stevens, Robert et al., "Ontology-Based Knowledge Representation for Bioinformatics", Briefings in Bioinformatics, 1, 4, 2000, 398-414

Steyerberg, E.W et al., "Application of Shrinkage Techniques in Logistic Regression Analysis: A Case Study", Statistica Neerlandica, 55(1), 2001, 76-88.

Strom, C. et al., "Three births after preimplantation genetic diagnosis for cystic fibrosis with sequential first and second polar body analysis", American Journal of Obstetrics and Gynecology, 178 (6), 1998, 1298-1306.

### US 11,519,035 B2

Page 17

### (56) References Cited

#### OTHER PUBLICATIONS

Strom, Charles M. et al., "Neonatal Outcome of Preimplantation Genetic Diagnosis by Polar Body Removal: The First 109 Infants", Pediatrics, 106(4), 2000, 650-653.

Stroun, Maurice et al., "Prehistory of the Notion of Circulating Nucleic Acids in Plasma/Serum (CNAPS): Birth of a Hypothesis", Ann. N.Y. Acad. Sci., 1075, 2006, 10-20.

Su, S.Y. et al., ""Inferring combined CNV/SNP haplotypes from genotype data"", Bioinformatics, vol. 26, No. 11,1, Jun. 1, 2010, 1437-1445.

Sun, Guihua et al., "SNPs in human miRNA genes affect biogenesis and function", RNA, 15(9), 2009, 1640-1651.

Sweet-Kind Singer, J. A. et al., "Log-penalized linear regression", IEEE International Symposium on Information Theory, 2003. Proceedings, 2003, 286.

Takano, T. et al., "Epidermal Growth Factor Receptor Gene Mutations and Increased Copy Numbers Predict Gefitinib Sensitivity in Patients With Recurrent Non-Small-Cell Lung Cancer", Journal of Clinical Oncology, vol. 23, No. 28, Oct. 1, 2005, 6829-6837.

Takara Biomedicals, , "Competitive PCR Guide", Lit. # L0126, Aug. 1999, 9 pages.

Taliun, D. et al., "Efficient haplotype block recognition of very long and dense genetic sequences", BMC Bioinformatics, vol. 15 (10), 2014, 1-18.

Tamura, et al., "Sibling Incest and formulation of paternity probability: case report", Legal Medicine, 2000, vol. 2, p. 189-196.

Tang, et al., , Multiplex fluorescent PCR for noninvasive prenatal detection of fetal-derived paternally inherited diseases using circulatory fetal DNA in maternal plasma, Eur J Obstet Gynecol Reprod Biol, 2009, v.144, No. 1, p. 35-39.

Tang, N. et al., "Detection of Fetal-Derived Paternally Inherited X-Chromosome Polymorphisms in Maternal Plasma", Clinical Chemistry, 45 (11), 1999, 2033-2035.

Tebbutt, S. J. et al., "Microarray genotyping resource to determine population stratification in genetic association studies of complex disease", BioTechniques, vol. 37, Dec. 2004, 977-985.

Ten Bosch, J., "Keeping Up With the Next Generation Massively Parallel Sequencing in Clinical Diagnostics", Journal of Molecular Diagnostics, vol. 10, No. 6, 2008, 484-492.

Tewhey, R. et al., "The importance of phase information for human genomics", Nature Reviews Genetics, vol. 12, No. 3, Mar. 1, 2011, 215-223.

The International Hapmap Consort, "The International HapMap Project", Nature, Vol. 426, Dec. 18, 2003, 789-796.

Thermofisher Scientific, "Ion AmpliSeq Cancer Hotspot Panel v2", Retrieved from the Internet: https://tools.thermofisher.com/content/sfs/brochures/Ion-AmpliSeq-Cancer-Hotspot-Panel-Flyer.pdf, 2015, 2 pages.

Thomas, M.R et al., "The Time of Appearance and Disappearance of Fetal DNA from the Maternal Circulation", Prenatal Diagnosis, 15, 1995, 641-646.

Tiersch, T. R. et al., "Reference Standards for Flow Cytometry and Application in Comparative Studies of Nuclear DNA Content", Cytometry, vol. 10, Mar. 21, 1989, 706-710.

Tong, Yu et al., "Noninvasive Prenatal Detection of Fetal Trisomy 18 by Epigenetic Allelic Ratio Analysis in Maternal Plasma: Theoretical and Empirical Considerations", Clinical Chemistry, 52(12), 2006, 2194-2202.

Tong, Yu K. et al., "Noninvasive Prenatal Detection of Trisomy 21 by Epigenetic-Genetic Chromosome-Dosage Approach", Clinical Chemistry, 56(1), 2010, 90-98.

Tounta, G. et al., "A Multiplex PCR for Non-invasive Fetal RHD Genotyping Using Cell-free Fetal DNA", in vivo, vol. 25, 2011, 411-418

Troyanskaya, Olga G. et al., "A Bayesian Framework for Combining Heterogeneous Data Sources for Gene Function Prediction (in *Saccharomyces cerevisiae*)", PNAS, 100(14), 2003, 8348-8353.

Tsui, Nancy B.Y et al., "Non-Invasive Prenatal Detection of Fetal Trisomy 18 by RNA-SNP Allelic Ratio Analysis Using Maternal Plasma SERPINB2 mRNA: A Feasibility Study", Prenatal Diagnosis, 29, 2009, 1031-1037.

Tu, J. et al., "Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis", BMC Genomics, vol. 13, No. 43, Jan. 25, 2012, 1-9.

Turner, E. et al., "Massively Parallel Exon Capture and Library-Free Resequencing Across 16 Genomes", Nature Methods, 6 (5), 2009, 315-316.

Tynan, J. A. et al., "Restriction Enzyme-Mediated Enhanced Detection of Circulating Cell-Free Fetal DNA in Maternal Plasma", The Journal of Molecular Diagnostics, vol. 13, No. 4, Jul. 2011, 382-389

Tzimagiorgis, G. et al., "Recovering circulating extracellular or cell-free RNA from bodily fluids", Cancer Epidemiology, vol. 35, 2011, 580-589.

Vallone, P. M. et al., "A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome", Int J Legal Medicine, vol. 118, Feb. 4, 2004, 147-157.

Vallone, Peter , "AutoDimer: A Screening Tool for Primer-Dimer and Hairpin Structures", Bio Techniques, 37, 2004, 226-231. Van Den Oever, J. M. et al., "Single Molecule Sequencing of Free

Van Den Oever, J. M. et al., "Single Molecule Sequencing of Free DNA from Maternal Plasma for Noninvasive Trisomy 21 Detection", Clinical Chemistry, vol. 58, No. 4, 2012, 699-706.

Varley, Katherine Elena et al., "Nested Patch PCR Enables Highly Multiplexed Mutation Discovery in Candidate Genes", Genome Res., 18(11), 2008, 1844-1850.

Verlinsky, Y. et al., "Over a Decade of Experience with Preimplanta-

Verlinsky, Y. et al., "Over a Decade of Experience with Preimplantation Genetic Diagnosis", Fertility and Sterility, 82 (2), 2004, 302-303.

Wagner, Jasenka et al., "Non-Invasive Prenatal Paternity Testing from Maternal Blood", Int. J. Legal Med., 123, 2009, 75-79. Wang, D. G. et al., "Large-Scale Identification, Mapping, and

Wang, D. G. et al., "Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome", Science, vol. 280, May 15, 1998, 1077-1082.

Wang, Eric et al., "Gestational Age and Maternal Weight Effects on Fetal Cell-Free DNA in Maternal Plasma", Prenatal Diagnosis, 33, 2013, 662-666.

Wang, Hui-Yun et al., "A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome", Genome Res., 15, 2005, 276-283.

Wang, T.L. et al., "Digital karyotyping", PNAS, vol. 99, No. 25, Dec. 10, 2002, 16156-16161.

Wang, W.-P. et al., "Multiplex single nucleotide polymorphism genotyping by adapter ligation-mediated allele-specific amplification". Analytical Biochemistry, vol. 355, May 5, 2006, 240-248.

Wang, Yuker et al., "Allele quantification using molecular inversion probes (MIP)", Nucleic Acids Research, vol. 33, No. 21, Nov. 28, 2005, 14 pgs.

Wapner, R. et al., "Chromosomal Microarray Versus Karyotyping for Prenatal Diagnosis", The New England Journal of Medicine, 367 (23), 2012, 2175-2184.

Wapner, R. et al., "First-Trimester Screening for Trisomies 21 and 18", The New England Journal of Medicine, vol. 349, No. 15, Oct. 9, 2003, 1405-1413.

Wapner, R. J. et al., "Expanding the scope of noninvasive prenatal testing: detection of fetal microdeletion syndromes", American Journal of Obstetrics & Gynecology, vol. 212, Dec. 17, 2014,

Watkins, N. et al., "Thermodynamic contributions of single internal rA •dA, rC • dC, rG • dG and rU • dT mismatches in RNA/DNA duplexes", Nucleic Acids Research, 9 (5), 2010, 1894-1902.

Weiss, C. A., "Chapter 8: Confidence Intervals for One Population Mean", Introductory Statistics, Sixth Edition, 2002, 340-381.

Wells, D, "Microarray for Analysis and Diagnosis of Human Embryos", 12th International Congress on Prenatal Diagnosis and Therapy Budapest Hungary 2004 9-17

Therapy, Budapest, Hungary, 2004, 9-17.
Wells, Dagan, "Advances in Preimplantation Genetic Diagnosis",
European Journal of Obstetrics and Gynecology and Reproductive
Biology, 115S, 2004, S97-S101.

### US 11,519,035 B2

Page 18

### (56) References Cited

#### OTHER PUBLICATIONS

Wells, Dagan, "Detailed Chromosomal and Molecular Genetic Analysis of Single Cells by Whole Genome Amplification and Comparative Genomic Hybridisation", Nucleic Acids Research, 27, 4, 1999, 1214-1218.

Wen, Daxing et al., "Universal Multiples PCR: A Novel Method of Simultaneous Amplification of Multiple DNA Fragments", Plant Methods, 8(32), Null, 2012, 1-9.

Widlak, P. et al., "Cleavage Preferences of the Apoptotic Endonuclease DFF 40 (Caspase-activated DNase or Nuclease) on Naked DNA and Chromatin Substrates", The Journal of Biological Chemistry, vol. 275, No. 11, Mar. 17, 2000, 8228-8232.

Wikipedia, "Buffy coat", Retrieved from "https://en.wikipedia.orgJw/index.php?title=Buffy\_coat&oldid=900992886", Jun. 9, 2019, 2 pgs.

Wikipedia, "Maximum a posteriori estimation", https://en.wikipedia.org/w/index.php?title=Maximum\_aposteriori\_estimation&oldid=26878808, [retrieved on Aug. 1, 2017], Oct. 30, 2005, 2 pages.

Wikipedia, "Stimulant", (available at https://en.wikipedia.org/wiki/Stimulant, accessed Mar. 14, 2016), 17 pages.

Wilton, et al., "Birth of a Healthy Infant After Preimplantation Confirmation of Euploidy by Comparative Genomic Hybridization", N. Engl. J. Med., 345(21), 2001, 1537-1541.

Wilton, L., "Preimplantation Genetic Diagnosis and Chromosome Analysis of Blastomeres Using Comparative Genomic Hybridization", Human Reproduction Update, 11 (1), 2005, 33-41.

Wittwer, C. T. et al., "Real-Time Multiplex PCR Assays", Methods, vol. 25, 2001, 430-448.

Wong, K. K. et al., "Allelic imbalance analysis by high-density single nucleotide polymorphic allele (SNP) array with whole genome amplified DNA", Nucleic Acids Research, vol. 32, No. 9, May 17, 2004, 8 pages.

Wright, C. et al., "The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis", Human Reproduction Update, vol. 15, No. 1, 2009, 139-151.

Wright, C. F. et al., "Cell-free fetal DNA and RNA in maternal

Wright, C. F. et al., "Cell-free tetal DNA and RNA in maternal blood: implications for safer antenatal testing", BMJ, vol. 39, Jul. 18, 2009, 161-165.

Wu, Y. Y. et al., "Rapid and/or high-throughput genotyping for

Wu, Y. Y. et al., "Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications". Clinica Chimica Acta, vol. 363, 2006, 165-176.

applications", Clinica Chimica Acta, vol. 363, 2006, 165-176. Xia, Tianbing et al., "Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs", Biochemistry, 37, 1998, 14719-14735. Xu, N. et al., "A Mutation in the Fibroblast Growth Factor Receptor 1 Gene Causes Fully Penetrant Normosmic Isolated Hypogonadotropic Hypogonadism", The Journal of Clinical Endocrinology & Metabo-

lism, vol. 92, No. 3, 2007, 1155-1158. Xu, S. et al., "Circulating tumor DNA identified by targeted sequencing in advanced-stage non-small cell lung cancer patients", Cancer Letters, vol. 370, 2016, 324-331.

Cancer Letters, vol. 370, 2016, 324-331.
Yamada, T. et al., "PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome", Nucleic Acids Research, vol. 34, 2006, W665-W669.

Acids Research, vol. 34, 2006, W665-W669.
Yeh, Iwei et al., "Knowledge Acquisition, Consistency Checking and Concurrency Control for Gene Ontology (GO)", Bioinformatics, 19, 2, 2003, 241-248.
You, Frank M. et al., "BatchPrimer3: A high throughput web

You, Frank M. et al., "BatchPrimer3: A high throughput web application for PCR and Sequencing Primer Design", BMC Bioinformatics, Biomed Central, London, GB, vol. 9, No. 1, May 29, 2008 (May 29, 2008), p. 253.

Yuan, X. et al., "Probability Theory-based SNP Association Study Method for Identifying Susceptibility Loci and Genetic Disease Models in Human Case-Control Data", IEEE Trans Nanobioscience, vol. 9, No. 4, Dec. 2010, 232-241.

Yung, T. K. et al., "Single-Molecule Detection of Epidermal Growth Factor Receptor Mutations in Plasma by Microfluidics Digital PCR in Non-Small Cell Lung Cancer Patients", Clinical Cancer Research, vol. 15, Mar. 10, 2009, 2076-2084.

Zachariah, R. et al., "Circulating cell-free DNA as a potential biomarker for minimal and mild endometriosis", Reproductive BioMedicine Online, vol. 18, No. 3, Jan. 27, 2009, 4007-411.

Zhang, J. et al., "Presence of Donor-and Recipient-derived DNA in Cell-free Urine Samples of Renal Transplantation Recipients: Urinary DNA Chimerism", Clinical Chemistry, vol. 45, No. 10, 1999, 1741-1746.

Zhang, L. et al., "Whole genome amplification from a single cell: Implications for genetic analysis", Proc. Nat'l. Acad. Sci. USA, vol. 89, Jul. 1992, 5847-5851.

Zhang, Rui et al., "Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing", Nature Methods, 11(1), 2014, 51-56.

Zhao, Xiaojun. et al., "An Integrated View of Copy Number and Allelic Alterations in the Cancer Genome Using Single Nucleotide Polymorphism Arrays", Cancer Research, 64, 2004, 3060-3071.

Zhong, X. et al., "Risk free simultaneous prenatal identification of fetal Rhesus D status and sex by multiplex real-time PCR using cell free fetal DNA in maternal plasma", Swiss Medical Weekly, vol. 131, Mar. 2001, 70-74.

Zhou, W. et al., "Counting Alleles Reveals a Connection Between Chromosome 18q Loss and Vascular Invasion", Nature Biotechnology, 19, 2001, 78-81.

Zimmermann, et al., "Noninvasive Prenatal Aneuploidy Testing of Chromosomes 13, 18, 21 X, and Y, Using targeted Sequencing of Polymorphic Loci", Prenatal Diagnosis, 32, 2012, 1-9. Zimmermann, B., "Declaration Under 37 CFR 1.32", filed in U.S.

Zimmermann, B., "Declaration Under 37 CFR 1.32", filed in U.S Appl. No. 14/171,587, Feb. 3, 2014, 4 pgs.

Zimmermann, B., "Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci, Supplemental Information", Prenatal Diagnosis, vol. 32, 2012, 7 pages.

Avgidou, K., et al., "Prospective first-trimester screening for trisomy 21 in 30,564 pregnancies", American Journal of Obstetrics and Gynecology, 2005, 1761-1767.

Barski, A., et al., "High-Resolution Profiling of Histone Methylations in the Human Genome", Cell, May 18, 2007, 823-837.

Bauer, M., et al., "A prospective analysis of cell-free fetal DNA concentration in maternal plasma as an indicator for adverse pregnancy outcome", Prenatal Diagnosis, 2006, 831-836.

Baxter, L. L., et al., "Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse", Human Molecular Genetics, Jan. 2000, 195-202.

Bennett, S. T., et al., "Toward the \$1000 human genome", Pharmacogenomics, 2005, 373-382.

Blow, N. , "The personal side of genomics", Nature, Oct. 4, 2007,  $627\text{-}630\,.$ 

Chiu, R.W.K., et al., "Hypermethylation of RASSF1A in Human and Rhesus Placentas", The American Journal of Pathology, Mar. 2007, 941-950.

Falcon, O. , "Screening for trisomy 21 by fetal tricuspid regurgitation, nuchal translucency and maternal serum free b-hCG and PAPP-A at 11 + to 13 + 6 weeks", Ultrasound Obstet Gynecol, 2006, 151-155.

Fan, J.-B., et al., "Highly Parallel SNP Genotyping", Cold Spring Harbor Symposia on Quantitative Biology, Feb. 2003, 69-78. Gautier, E., et al., "Fetal RhD genotyping by maternal serum

Gautier, E., et al., "Fetal RhD genotyping by maternal serum analysis: A two-year experience", American Journal of Obstetrics and Gynecology, 2005, 666-669.

Illumina , "Automated GoldenGate $^{TM}$  Genotyping on the BeadStation 500", Pub. No. 970-2004-002, 2004, 2 pages.

Illumina, "GoldenGate Assay Workflow: Illumina's GoldenGate assay protocol provides high-quality, high-multiplex genotyping results with a streamlined workflow", Pub. No. 370-2004-006, 2004, 2 pages.

2004, 2 pages.

Illumina , "Illumina Extends BeadArray Technology to Address Wider Range of SNP Genotyping Projects; New Microarray Offerings Enable Genotyping at 384 and 786 Multiplex", https://www.businesswire.com/news/home/20040504006011/en/Illumina-Extends-BeadArray-Technology-to-Address-Wider-Range-of-SNP-Genotyping-Projects-New-Microarray-Offerings-Enable-Genotyping-at-384-and-786-Multiplex, May 4, 2004, 2 pages.

## US 11,519,035 B2

Page 19

### (56) References Cited

#### OTHER PUBLICATIONS

Illumina, "Illumina® Beadstation 500: A Scalable System That Grows With Your Research Requirements", Pub. No. 970-2005-003, Jul. 1, 2005, 4 pages.

Illumina, "Illumina Announces Benchtop SNP Genotyping System", Press Release, Nov. 5, 2003, 3 pages.

Illumina, "Illumina Begins Shipment of BeadStation 500G Benchtop Genotyping System", Press Release, Apr. 15, 2004, 3 pages. Illumina, "Preparing Samples for Sequencing Genomic DNA", Part

# 11251892 Rev. A, 2007, 18 pages.

Illumina , "Products & Services", support contact sitemap legal privacy +1 858.202.4566 ©2007 Illumlna, Inc. All rights reserved. https://we b. archive .o rg/web/20070321 001 025/http://www. ii lu m ina .co m/pagesn m. ii mn?ID= 70, Mar. 21, 2007, 3 pages.

Illumina , "Technology: Solexa Sequencing Technology", https://wwb.archive.org/web/20070521 081517 /http://www.Illumina.com/pages .ilmn?I D=203, May 21, 2007, 1 page.

Jett, K. , et al., "Clinical and genetic aspects of neurofibromatosis

Jett, K., et al., "Clinical and genetic aspects of neurofibromatosi 1", Genetics in Medicine, Jan. 2010, 11 pages.

Johnson, D. S., et al., "Genome-Wide Mapping of in Vivo Protein—DNA Interactions", Science, Jun. 8, 2007, 1497-1502.

Kamel, A. M., et al., "A simple strategy for breakpoint fragment determination in chronic myeloid leukemia", Cancer Genetics and Cytogenetics, 2000, 110-115.

Lapaire, O., et al., "Array-CGH analysis of cell-free fetal DNA in

Lapaire, O., et al., "Array-CGH analysis of cell-free fetal DNA in 10 mL of amniotic fluid supernatant", Prenatal Diagnosis, May 17, 2007, 616-621.

Lapierre, J.M., et al., "Analysis of uncultured amniocytes by comparative genomic hybridization: a prospective prenatal study", Prenatal Diagnosis, 2000, 123-131.

Lasken, R. S., et al., "Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens", Trends in Biotechnology, Dec. 2003, 531-535.

Li, Y., et al., "Detection of Paternally Inherited Fetal Point Mutations for b-Thalassemia Using Size-Fractionated Cell-Free DNA Maternal Plasma" IAMA Apr. 13, 2005, 843-849

DNA Maternal Plasma", JAMA, Apr. 13, 2005, 843-849. Lo, Y.M.D., "Fetal DNA in Maternal Plasma: Biology and Diagnostic Applications", Clinical Chemistry, 2000, 1903-1906. Masuzaki, H., et al., "Detection of cell free placental DNA in maternal plasma: direct evidence from three cases of confined placental massicism". J.Med. Genet. 2004, 289-292.

placental mosaicism", J Med Genet, 2004, 289-292. Nagalla, S. R., et al., "Proteomic Analysis of Maternal Serum in Down Syndrome: Identification of Novel Protein Biomarkers", Journal of Proteome Research, Mar. 21, 2007, 1245-1257.

Oliphant, A., et al., "Bead.Array™ Technology: Enabling an Accurate, Cost-Effective Approach to High-Throughput Genotyping". Bio Techniques. Jun 2002, S56-S6.

ing", Bio Techniques, Jun 2002, S56-S6.

Pask, R., et al., "Investigating the utility of combining 29 whole genome amplification and highly multiplexed single iucleotide polymorphism BeadArray TM genotyping", BMC Biotechnology, Jul. 27, 2004, 8 pages.

Patil, N., et al., "Blocks of Limited Haplotype Diversity Revealed

Patil, N., et al., "Blocks of Limited Haplotype Diversity Revealed by High-Resolution Scanning of Human chromosome 21", Science, Nov. 23, 2001, 1719-1723.

Paunio, T., et al., "Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA", Clinical Chemistry, 1996, 1382-1390.

Philip, J., et al., "Late First-Trimester Invasive Prenatal Diagnosis: Results of an International Randomized Trial", American College of Obstetricians and Gynecologists, Jun. 2004, 1164-1173.

Robertson, G., et al., "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing", Nature Methods, Aug. 2007, 651-657. Roman, B. L., et al., "Non-Radioisotopic AFLP Method Using PCR

Roman, B. L., et al., "Non-Radioisotopic AFLP Method Using PCR Primers Fluorescently Labeled with CyA 5", BioTechniques, Feb. 1999, 236-238.

Ruano, G., et al., "Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules", Proc. Nati. Acad. Sci. USA, Aug. 1990, 6296-6300. Samura, O., et al., "Diagnosis of Trisomy 21 in Fetal Nucleated Erythrocytes from Maternal Blood by Use of Short Tandem Repeat Sequences", Clinical Chemistry, 2001, 1622-1626.

Seppo, A., et al., "Detection of circulating fetal cells utilizing automated microscopy: potential for noninvasive prenatal diagnosis of chromosomal aneuploidies", Prenatal Diagnosis, Jul. 22, 2008, 815-821.

Spencer, K., et al., "Maternal serum levels of dimeric inhibin A in pregnancies affected by trisomy 21 in the first trimester", Prenatal Diagnosis, 2001, 441-444.

Spencer, K., et al., "Maternal serum levels of total activin-A in first-trimester trisomy 21 pregnancies", Prenatal Diagnosis, 2001, 270-273.

Swinkels, D. W., et al., "Effects of Blood-Processing Protocols on Cell-free DNA Quantification in Plasma", Clinical Chemistry, 2003, 525-526.

Tsangaris, G. T., et al., "Proteomic analysis of amniotic fluid in pregnancies with Down syndrome", Proteomics, 2006, 4410-4419. Vogelstein, B., et al., "Digital PCR", Proc. Natl. Acad. Sci. USA, Aug. 1999, 9236-9241.

Zimmermann, B., et al., "Digital PCR: a powerful new tool for noninvasive prenatal diagnosis?", Prenatal Diagnosis, Nov. 10, 2008, 1087-1093.

Zimmermann, B., et al., "Novel Real-Time Quantitative PCR Test for Trisomy 21", Clinical Chemistry, 2002, 362-363.

Zimmermann, B., et al., "Optimized Real-Time Quantitative PCR Measurement of Male Fetal DNA in Maternal Plasma", Clinical Chemistry, 2005, 1598-1604.

Zimmermann, B., et al., "Real-Time Quantitative Polymerase Chain Reaction Measurement of Male Fetal DNA in Maternal Plasma", Methods in Molecular Medicine, 2007, 43-49.

Zimmermann, B., et al., "Use of Real-Time Polymerase Chain Reaction for the Detection of Fetal Aneuploidies", Methods in Molecular Biology, Feb. 2006, 83-100.

Canick, J. A., et al., "The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies", Prenatal Diagnosis, 2013, 667-674.

Chitty, L. S., et al., "Noninvasive Prenatal Screening for Genetic Diseases Using Massively Parallel Sequencing of Maternal Plasma DNA", Cold Spring Harbor Perspectives in Medicine, 2015, 20

Ding, C., et al., "MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis", PNAS, Jul. 20, 2004, 10762-10767.

Landegren, U., et al., "Padlock and proximity probes for in situ and array-based analyses: tools for the post-genomic era", Comparative and Functional Genomics. 2003. 525-530.

and Functional Genomics, 2003, 525-530.

Matsuzaki, H., et al., "Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays" Nature Methods, Nov. 2004, 109-111

oligonucleotide arrays", Nature Methods, Nov. 2004, 109-111. Morris, J. K., et al., "Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register", BMJ Online, Oct. 2009, 5 pages.

Nilsson, M., et al., "Padlock Probes: Circularizing Oligonucle-

Nilsson, M., et al., "Padlock Probes: Circularizing Oligonucleotides for Localized DNA Detection", Science, Sep. 10, 1994, 2085-2088.

Parameswaran, P., et al., "A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing", Nucleic Acids Research, Oct. 11, 2007, 9 pages.

Syvanen, A.C., "Toward genome-wide SNP genotyping", Nature Genetics Supplement, Jun. 2005, S5-S10.

Wu, T.L., et al., "Cell-free DNA: measurement in various carcinomas and establishment of normal reference range", Clinica Chimica Acta, 2002, 77-87.

Yaron, Y., "The implications of non-invasive prenatal testing failures: a review of an under-discussed phenomenon", Prenatal Diagnosis, 2016, 391-396.

"Abstracts for CNAPS III Circulating Nucleic Acids in Plasma and Serum and Serum Proteomics", Clinical Chemistry, vol. 49, No. 11, 2003, 33 pages.

"Abstracts for CNAPS IV Circulating Nucleic Acids in Plasma/ Serum", Fourth International Conference on Circulating Nucleic Acids in Plasma/Serum (CNAPS-IV), 2005, 40 pages. Case: 24-1324 Document: 42-1 Page: 55 Filed: 03/18/2024

### US 11,519,035 B2

Page 20

#### (56)References Cited

#### OTHER PUBLICATIONS

Ambardar, S. et al., "High Throughput Sequencing: An Overview of Sequencing Chemistry", Indian J. Microbiol., vol. 56, No. 4, 2016,

Anker, P. et al., "The Second International Symposium on Circulating Nucleic Acids in Plasma and Serum (CNAPS-2) held in conjunction with the 6th Annual Scientific Symposium of the Hong Kong Cancer Institute", Clinical Chemistry, vol. 47, No. 2, 2001,

Bale, J. R. et al., "Reducing Birth Defects: Meeting the Challenge in the Developing World", Institute of Medicine of the National Academies, 2003, 270 pgs

Choi, Y. et al., "Comparison of phasing strategies for whole human genomes", PLOS Genetics, Apr. 5, 2018, 26 pages.

Clausen, F. B. et al., "Improvement in fetal DNA extraction from maternal plasma. Evaluation of the NucliSens Magnetic Extraction system and the QIAamp DSP Virus Kit in comparison with the QIAamp DNA Blood Mini Kit", Prenatal Diagnosis, vol. 27, 2007, 6 - 10.

Di, X. et al., "Dynamic model based algorithms for screening and genotyping", Bioinformatics, vol. 21, No. 9, 2005, 1958-1963. Dressman, D. et al., "Transforming single DNA molecules into

fluorescent magnetic particles for detection and enumeration of genetic variations", PNAS, vol. 100, No. 15, Jul. 22, 2003, 8817-

Eltoukhy, H. et al., "Modeling and Base-Calling for DNA Sequencing-By-Synthesis", IEEE, 2006, II-1032-II-1035.

Erlich, R. L. et al., "Next-generation sequencing for HLA typing of class loci", BMC Genomics, vol. 12, No. 42, 2011, 13 pages

Fortina, P. et al., "DOP-PCR Amplification of Whole Genomic DNA and Microchip-Based Capillary Electrophoresis", Methods in Molecular Biology: Capillary Electrophoresis of Nucleic Acids, vol. II Practical Applications of Capillary Electrophoresis, 2001, 211-

Griffiths, A. J. et al., "An Introduction to Genetic Analysis", Sixth

Edition, 1996, 5 pages.
Grunenwald, H., "Optimization of Polymerase Chain Reactions", Methods in Biology, vol. 226, 2003, 89-99

Hosono, S. et al., "Unbiased Whole-Genome Amplification Directly From Clinical Samples", Genome Research, vol. 13, 2003, 954-964. Huang, D. J. et al., "Use of an Automated Method Improves the Yield and Quality of Cell-Free Fetal DNA Extracted from Maternal Plasma", Clinical Chemistry, vol. 51, No. 12, 2005, 2419-2420.

Illumina, "History of Sequencing by Synthesis", https://www. illumina.com/science/technology/next-generation-sequencing/illumina-

sequencing-history.html, 2020, 3 pages.
Illumina, "Preparing Samples for Sequencing Genomic DNA". (available at http://zazil.ibt.unam.mx/usmb/wpcontent/uploads/2016/ 05/1003806\_Genomic\_DNA\_Sample\_Prep.pdf), Part #1003806 Rev. A, 2007, 20 pages.

Innan, H. et al., "The Pattern of Polymorphism on Human Chromosome 21", Genome Research, vol. 13, 2003, 1158-1168. Jiang, P. et al., "The Long and Short of Circulating Cell-Free DNA

and the Ins and Outs of Molecular Diagnostics", Trends in Genetics, vol. 32, No. 6, Jun. 2016, 360-371.

Kirkness, E. F. et al., "Sequencing of isolated sperm cells for direct haplotyping of a human genome", Genome Research, vol. 23, 2013, 826-832

Lu, S. et al., "Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing", Science, vol. 338, Dec. 21, 2012, 1627-1630.

Meyerson, M. et al., "Advances in understanding cancer genomes through second-generation sequencing", Nature Reviews: Genetics, vol. 11, Oct. 2010, 685-696.

Mikkelsen, T. S. et al., "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells", Nature, vol. 448, No. 2, Aug. 2007, 553-562.

Nishigaki, K. et al., "Random PCR-Based Genome Sequencing: A Non-Divide-and-Conquer Strategy", DNA Research, vol. 7, 2000,

Pont-Kingdon, G. et al., "Rapid Detection of Aneuploidy (Trisomy 21) by Allele Quantification Combined with Melting Curves Analysis of Single-Nucleotide Polymorphism Loci", Clinical Chemistry, vol. 49, No. 7, 2003, 1087-1094.

Quinlan, M. P., "Amniocentesis: Indications and Risks", American Medical Association Journal of Ethics: Virtual Mentor, vol. 10, No. 5, May 2008, 304-306.

Reeves, R. H. et al., "Too much of a good thing: mechanisms of gene action in Down syndrome", Trends in Genetics, vol. 17, No. 2, Feb. 2, 2001, 83-88.

Rhoads, A. et al., "PacBio Sequencing and Its Applications", Genomics Proteomics Bioinformatics, vol. 13, Nov. 2, 2015, 278-

Schubert, "Picking out prenatal DNA", Nature Medicine, vol. 10, No. 785, Aug. 2004, 1 page.

Shendure, J. et al., "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome", Science, Nov. 30, 2007, 18-24

Shendure, J. et al., "Next-generation DNA sequencing", Nature Biotechnology, vol. 26, No. 10, Oct. 2008, 1135-1145.

Shokralla, S. et al., "Next-generation DNA barcoding: using nextgeneration sequencing to enhance and accelerate DNA barcode capture from single specimens", Molecular Ecology Resources, vol. 14, 2014, 892-901.

Sivertsson, A. et al., "Pyrosequencing as an Alternative to Single-Strand Conformation Polymorphism Analysis for Detection of N-ras Mutations in Human Melanoma Metastases", Clinical Chemistry, vol. 48, No. 12, 2002, 2164-2170.

Stewart, C. M. et al., "Circulating cell-free DNA for non-invasive cancer management", Cancer Genetics, vol. 228-229, 2018, 169-

Taback, B. et al., "Quantification of Circulating DNA in the Plasma and Serum of Cancer Patients", Ann. N.Y. Acad. Sci, vol. 1022,

Von Eggeling, F. et al., "Applications of Random PCR", Cellular and Molecular Biology, vol. 41, No. 5, 1995, 653-670.

Wang, J. et al., "Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm", Cell, vol. 150, Jul. 20, 2012, 402-412.

Wei, C. et al., "Detection and Quantification by Homogeneous PCR of Cell-free Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 47, No. 2, 2001, 336-338.

Winsor, E. J. et al., "Maternal Cell Contamination in Uncultured Amniotic Fluid", Prenatal Diagnosis, vol. 16, 1996, 49-54.

Zheng, S. et al., "Whole Genome Amplification Increases the Efficiency and Validity of Buccal Cell Genotyping in Pediatric Populations1", Cancer Epidemiology, Biomarkers & Prevention, vol. 10, Jun. 2001, 697-700.

Zlotogora, J., "Penetrance and expressivity in the molecular age", Genetics in Medicine, vol. 5, No. 5, 2003, 347-352. Agbor-Enoh, S. et al., "Donor-derived cell-free DNA predicts

allograft failure and mortality after lung transplantation", EBioMedicine, vol. 40, 2019, 541-553. Kanou, et al., "Cell-free DNA in human ex vivo lung perfusate as

a potential biomarker to predict the risk of primary graft dysfunction in lung transplantation", The Journal of Heart and Lung Transplantation, vol. 36, No. 45, 2017, S187.

Marshutina, N. V. et al., "Comparative Clinical and Diagnostic

Significance of Some Serological Tumor Associated Markers for Different Histological Types of Lung Cancer", Russian Oncological Journal, vol. 3, 2010, 13-16.

Xu, W. et al., "A Novel Universal Primer-Multiplex-PCR Method with Sequencing Gel Electrophoresis Analysis", PLOS One, vol. 7, No. 1, Jan. 17, 2012, 10 pgs.

Couraud, S. et al., "Noninvasive Diagnosis of Actionable Mutations by Deep Sequencing of Circulating Free DNA in lung Cancer from Never-Smokers: A Proof-of-Concept Study from BioCAST/IFCT-1002", Clinical Cancer Research, vol. 20, No. 17, Jul. 10, 2014, 4613-4624

Couraud, S. et al., "Supplementary Data for Noninvasive Diagnosis of Actionable Mutations by Deep Sequencing of Circulating Free DNA in lung Cancer from Never-Smokers: A Proof-of-Concept Study from BioCAST/IFCT-1002", 2014, 13 pages.

Case: 24-1324 Document: 42-1 Page: 56 Filed: 03/18/2024

## US 11,519,035 B2

Page 21

#### References Cited (56)

#### OTHER PUBLICATIONS

Langmore, J., "Quality Control and Pre-Qualifications of NGS Libraries Made from Clinical Samples", ABRF 2013 Satellite Workshop, Mar. 2, 2013, 35 pages.

Takashima, Y. et al., "Expansion-contraction of photoresponsive artificial muscle regulated by host-guest interactions", Nature Communications, vol. 3, No. 1270, Dec. 11, 2012, 8 pages.

Abaan, O. D. et al., "The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology", Cancer Res., vol. 73, No. 14, Jul. 15, 2013, 4372-4382.

Amicucci, P. et al., "Prenatal Diagnosis of Myotonic Dystrophy Using Fetal DNA Obtained from Maternal Plasma", Clinical Chemistry, vol. 46, No. 2, 2000, 301-302.

Anker, P. et al., "Circulating DNA in Plasma or Serum", Medicina, vol. 60, 2000, 699-702.

Auld, D. S., "Use of Chelating Agents to Inhibit Enzymes", Methods in Enzymology, vol. 158, 1988, 110-114. Banfi, G et al., "The role of ethylenediamine tetraacetic acid

(EDTA) as in vitro anticoagulant for diagnostic purposes", Clin. Chem., vol. 45, No. 5, 2007, 565-576.

Barra, G. B. et al., "EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples", Clinical Biochemistry, vol. 48, 2015, 976-981.
Bergen, A. W. et al., "Effects of DNA mass on multiple displace-

ment whole genome amplification and genotyping performance", BMC Biotechnology, vol. 5, No. 24, Sep. 16, 2005, 11 pgs.

Bischoff, F. Z. et al., "Cell-free fetal DNA in maternal blood: kinetics, source and structure", Human Reproduction Update, vol. 11, No. 1, 2005, 59-67.

Bischoff, F. Z. et al., "Intact fetal cells in maternal plasma: are they really there?", Lancet, vol. 361, 2003, 139-140.

Board, R.E. et al., "Detection of BRAF mutations in the tumour and serum of patients enrolled in the AZD6244 (ARRY-142886) advanced melanoma phase II study", British Journal of Cancer, vol. 101, 2009, 1724-1730.

Boudsocq, F. et al., "Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archael DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poln", Nucleic Acids Research, vol. 29, No. 22, 2001, 4607-4616.

Bouma, B. N. et al., "Human Blood Coagulation Factor", The Journal of Biological Chemistry, vol. 252, No. 18, 1977, 6432-6437. Brinza, D. et al., "2SNP: scalable phasing based on 2-SNP haplotypes", Bioinformatics, vol. 22, No. 3, 2006, 371-373

Browning, S. R. et al., "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering", The American Journal of Human Genetics, vol. 81, Nov. 2007, 1084-1097.

Bryant, A. P., "Terminology of Sugars", Ind. Eng. Chem., vol. 26, No. 2, 1933, 231.

Burkey, B. F. et al., "Hepatic apolipoprotein J is secreted as a lipoprotein", Journal of Lipid Research, vol. 33, 1992, 1517-1526. Calin, G. A. et al., "A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia", N Engl J Med, vol. 353, 2005, 1793-1801. Cao, Y. et al., "Clinical Evaluation of Branched DNA Signal

Amplification for Quantifying HIV Type 1 in Human Plasma", AIDS Research and Human Retroviruses, vol. 11, No. 3, 1995,

Chen, C. P. et al., "Fetal DNA in maternal plasma: the prenatal detection of a paternally inherited fetal aneuploidy", Prenatal Diagnosis, vol. 20, 2000, 353-357.

Chim, S. S. et al., "Detection and Characterization of Placental MicroRNAs in Maternal Plasma", Clinical Chemistry, vol. 54, No. 3, 2008, 482-490.

Chinnapapagari, S. K. et al., "Treatment of Maternal Blood Samples with Formaldehyde Does Not Alter the Proportion of Circulatory Fetal Nucleic Acids (DNA and mRNA) in Maternal Plasma", Clinical Chemistry, vol. 51, No. 3, 2005, 653-655.

Chung, G. T. et al., "Lack of Dramatic Enrichment of Fetal DNA in Maternal Plasma by Formaldehyde Treatment", Clinical Chemistry, vol. 51, No. 3, 2005, 655-658.

Ciriello, G. et al., "Emerging landscape of oncogenic signatures across human cancers", Nature Genetics, vol. 45, No. 10, Oct. 2013, 1127-1135.

Delaneau, O. et al., "Shape-IT: new rapid and accurate algorithm for haplotype inference", BMC Bioinformatics, vol. 9, No. 540, Dec. 16, 2008, 14 pages

Dias-Santagata, D. et al., "BRAF V600E Mutations Are Common in Pleomorphic Xanthoastrocytoma: Diagnostic and Therapeutic Implications", PLoS One, vol. 6, No. 3, Mar. 2011, 9 pages

Dickover, R. E. et al., "Optimization of Specimen-Handling Procedures for Accurate Quantitation of Levels of Human Immunodeficiency Virus RNA in Plasma by Reverse Transcriptase PCR", Journal of Clinical Microbiology, vol. 36, No. 4, 1998, 1070-1073. Dowd, P. et al., "On the mechanism of the anticlotting action of vitamin R quinone", Proc. Natl. Acad. Sci. USA, vol. 92, 1995, 8171-8175

Downward, J., "Targeting RAS Signalling Pathways in Cancer Therapy", Nature Reviews, vol. 3, Jan. 2003, 11-22.

Fronen, L. et al., "HaploRec: efficient and accurate large-scale reconstruction of haplotypes", BMC Bioinformatics, vol. 7, No. 542, Dec. 22, 2006, 18 pages.

Fackenthal, J. D. et al., "Aberrant RNA splicing and its functional

consequences in cancer cells", Disease Models & Mechanisms, vol. 1, 2008, 37-42.

Fortina, P. et al., "Detection of the most common mutations causing beta-thalassemia in Mediterraneans using a multiplex amplification refractory mutation system (MARMS)", Genome Res., vol. 2, 1992,

Gu, H. et al., "Diagnostic role of microRNA expression profile in the serum of pregnant women with fetuses with neural tube defects", Journal of Neurochemistry, vol. 122, 2012, 641-649

Hahn, S. et al., "Current applications of single-cell PCR", CMLS Cellular and Molecular. Life Sciences, vol. 57, 2000, 96-105.

Howie, B. N. et al., "A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies", PLoS Genetics, vol. 5, No. 6, Jun. 2009, 15 pages.

Hu, Y. et al., "Detection of Extrahepatic Hepatitis C Virus Replication by a Novel, Highly Sensitive, Single-Tube Nested Polymerase Chain Reaction", Am. J. Clin Pathol., vol. 119, 2003, 95-100.

Huang, J. et al., "Whole genome DNA copy number changes identified by high density oligonucleotide arrays", Human Genomics, vol. 1, No. 4, May 2004, 287-299.

Hung, E.C.W. et al., "Detection of circulating fetal nucleic acids: a review of methods and applications", J. Clin. Pathol., vol. 62, 2009, 308-313.

Ivanov, M. et al., "Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation", BMC Genomics, vol. 16 (Suppl 13):S1, Jun. 2015, 12 pgs.

Jennings, C. et al., "Investigation of Effects of Acid Citrate Dextrose and EDTA on Ability to Quantitatively Culture Human Immuno-

deficiency Virus", Journal of Clinical Microbiology, vol. 38, No. 9, 2000, 3522.

Jewesburty, E.C.O., "Reactions after Transfusion of Stored Blood", The British Medical Journal, vol. 1, No. 4191, 1941, 664-665.

Johnson, J. B. et al., "Differential mechanisms of complementmediated neutralization of the closely related paramyxoviruses simian virus 5 and mumps virus", Virology, vol. 376, No. 1, 2008, 112-123. Johnson, K. L. et al., "Interlaboratory Comparison of Fetal Male DNA Detection from Common Maternal Plasma Samples by Real-Time PC", Clinical Chemistry, vol. 50, No. 3, 2004, 516-521.

Keith, L. et al., "Clinical Experience With the Prevention of Rh-Isoimmunization: A Historical Comparative Analysis", American Journal of Reproductive Immunology, vol. 5, 1984, 84-89.

Kiernan, J. A., "Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: What they are and what they do.", Microscopy Today, vol. 1, 2000, 8-12. Kimmel, G. et al., "GERBIL: Genotype resolution and block

identification using likelihood", PNAS, vol. 102, No. 1, Jan. 4, 2005, 158-162.

## US 11,519,035 B2

Page 22

#### (56) References Cited

#### OTHER PUBLICATIONS

Kohler, C. et al., "Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors", Molecular Cancer, vol. 8, No. 105, Nov. 17, 2009, 9 pages.

Kumar, P. et al., "Ethylenegycol-Bis-(B-Aminoethylether)Tetraacetate as a Blood Anticoagulant: Preservation of Antigen-Presenting Cell Function and Antigen-Specific Proliferative Response of Peripheral Blood Mononuclear Cells from Stored Blood", Clinical and Diagnostic Laboratory Immunology, vol. 7, No. 4, 2000, 578-583.

Lecomte, T. et al., "Detection of Free-Circulating Tumor-Associated DNA in Plasma of Colorectal Cancer Patients and Its Association With Prognosis", Int. J. Cancer, vol. 100, 2002, 542-548.

Lee, T. et al., "Down syndrome and cell-free fetal DNA in archived maternal serum", AmJ Obstet Gynecol, vol. 187, No. 5, 1217-1221, Nov. 2002.

Lee, T.H. et al., "Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma", Transfusion, vol. 41, Feb. 2001, 276-282. Lichtenstein, A. V. et al., "Circulating Nucleic Acids and Apopto-

Lichtenstein, A. V. et al., "Circulating Nucleic Acids and Apoptosis", Annals New York Academy of Sciences, vol. 945, 1993, 239-249.

Lo, Y.M.D. et al., "Prenatal diagnosis: progress through plasma nucleic acids", Nature Reviews, vol. 8, 2007, 71-77.

Lovmar, L. et al., "Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DN", Nucleic Acids Research, vol. 31, No. 21, 2003, 9 pgs.

Mackiewicz, D. et al., "Distribution of Recombination Hotspots in the Human Genome—A Comparison of Computer Simulations with Real Data", PLOS One, vol. 8, No. 6, Jun. 2013, 11 pages.

McDonald, J. P. et al., "Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs", Nucleic Acids Research, vol. 34, No. 4, 2006, 1102-1111. Murali, R. et al., "Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: The Fab is directed against an intermediate in the helix-coil dynamics of the enzyme", Proc. Natl. Acad. Sci. USA, vol. 95, Oct. 1998, 12562-12567.

Olive, M. et al., "Characterization of the DiFi Rectal Carcinoma Cell Line Derived from a Familial Adenomatous Polyposis Patient", In Vitro Cellular & Developmental Biology, vol. 29A, No. 3, Part 1, Mar. 1993, 239-248. Olney, R. S. et al., "Chorionic Villus Sampling and Amniocentesis:

Olney, R. S. et al., "Chorionic Villus Sampling and Amniocentesis: Recommendations for Prenatal Counseling", MMWR: Recommendations and Reports, 44(RR-9), Jul. 21, 1995, 1-12.

Parker, A. V. et al., "The Effect of Sodium Citrate on the Stimulation of Polymorphonuclear Leukocytes", Investigative Ophthalmology & Visual Science, vol. 26, 1985, 1257-1261

& Visual Science, vol. 26, 1985, 1257-1261.
Pelizzari, C. A. et al., "Quantitative analysis of DNA array autoradiographs", Nucleic Acids Research, vol. 28, No. 22, 2000, 4577-4581.

Qin, Z. S. et al., "Partition-Ligation—Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms", Am. J. Hum Genet., vol. 71, 2002, 1242-1247. Quan, P. C. et al., "Studies on the mechanism of NK cell lysis", The Journal of Immunology, vol. 128, 1982, 1786-1791.

Rosado, J. A. et al., "Tyrosine kinases activate store-mediated Ca2+entry in human platelets through the reorganization of the actin cytoskeleton", Biochem. J., vol. 351, 2000, 429-437.

Rosen, D. R. et al., "Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis", Nature, vol. 362, Mar. 4, 1993, 59-62.

Ryan, B. M. et al., "A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up", Gut, vol. 52, 2003, 101-108.

Sahukhal, G. S. et al., "msaABCR operon positively regulates biofilm development by repressing proteases and autolysis in *Staphlococcus aureus*", FEMS Microbiology Letters, vol. 362, No. 4, 2015, 1-10.

Saito, H. et al., "Prenatal DNA diagnosis of a single-gene disorder from maternal plasma", The Lancet, vol. 356, Sep. 30, 2000, 1170. Scheet, P. et al., "A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase", The American Journal of Human Genetics, vol. 78, Apr. 2006, 629-644.

Schwarzenbach, H. et al., "Evaluation of cell-free tumour DNA and RNA in patients with breast cancer and benign breast disease", Molecular BioSystems, vol. 7, 2011, 2848-2854.

Shinozaki, M. et al., "Utility of Circulating B-RAF DNA Mutation in Serum for Monitoring Melanoma Patients Receiving Biochemotherapy", Clin Cancer Res, vol. 13, No. 7, Apr. 1, 2007, 2068-2074.

Solomon, M. J. et al., "Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures", Proc. Natl. Acad. Sci. USA, vol. 82, 1985, 6470-6474.

Stephens, M. et al., "Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation", Am. J. Hum. Genet., vol. 76, 2005, 449-462.

Su, Z. et al., "A Platform for Rapid Detection of Multiple Oncogenic Mutations With Relevance to Targeted Therapy in Non-Small-Cell Lung Cancer", The Journal of Molecular Diagnostics,, vol. 13, No. 1, Jan. 2011, 74-84.

Thavarajah, R. et al., "Chemical and physical basics of routine formaldehyde fixation", Journal of Oral and Maxillofacial Pathology, vol. 16, No. 3, 2012, 400-405.
Tsui, N. B. et al., "Systematic micro-array based identification of

Tsui, N. B. et al., "Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling", J. Med. Genet, vol. 41, 2004, 461-467. Urbaniak, S. J. et al., "RhD haemolytic disease of the fetus and the newborn", Blood Reviews, vol. 14, 2000, 44-61.

Van Uitert, I. et al., "The influence of different membrane components on the electrical stability of bilayer lipid membranes", Biochimica et Biophysica Acta, vol. 1798, 2010, 21-31.

Wang, S. et al., "Potential Clinical Significance of a Plasma-Based KRAS Mutation Analysis in Patients with Advanced Non-Small Cell Lung Cancer", Clin Cancer Res, vol. 16, No. 4, Feb. 15, 2010, 1324-1330.

Yamada, T. et al., "Detection of K-ras Gene Mutations in Plasma DNA of Patients with Pancreatic Adenocarcinoma: Correlation with Clinicopathological Features", Clinical Cancer Research, vol. 4, Jun. 1998, 1527-1532.

Beck, J. , et al., "Profile of the Circulating DNA in Apparently Healthy Individuals", Clinical Chemistry, 2009, 730-738. Bianchi, D W, et al., "Insights Into Fetal and Neonatal Development

Bianchi, D W, et al., "Insights Into Fetal and Neonatal Development Through Analysis of Cell-Free RNA in Body Fluids", Early Human Development, vol. 86, No. 11, Nov. 2010, 747-752. Blomquist, T M, et al., "Targeted RNA-Sequencing with Competi-

Blomquist, T.M., et al., "Targeted RNA-Sequencing with Competitive Multiplex-PCR Amplicon Libraries", Plos One, vol. 8, Issue 11, Nov. 2013, 14 pages.

Broude, N. E., et al., "High Multiplexity PCR Based on PCR

Broude, N. E., et al., "High Multiplexity PCR Based on PCR Suppression", DNA Amplification Current Technologies and Applications, 2004, 61-76.

Broude, N E, et al., "High-Level Multiplex DNA Amplification", Antisense & Nucleic Acid Drug Development, vol. 11, 2001, 327-332.

Broude, N. E., et al., "Multiplex Allele-specific Target Amplification based on PCR Suppression", PNAS, vol. 98, No. 1, Jan. 2, 2001, 206-211.

Campbell, P. J., et al., "Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing", PNAS, Sep. 2, 2008, 13081-13096

Castleberry, C. D., et al., "Quantification of Circulating Cell—Free DNA in Pediatric Heart Transplant Recipients", Journal of Heart and Lung Transplantation, Apr. 1, 2011, S139.

Edwards, M. C., et al., "Multiplex PCR: Advantages, Development, and Applications", Genome Research, 1994, S65-S75.

Fan, C H, et al., "Detection of Aneuploidy with Digital Polymerase Chain Reaction", Analytical Chemistry, vol. 79, No. 19, Oct. 1, 2007, 7576-7579.

Frohman, M A, et al., "On Beyond Classic RACE (Rapid Amplification of cDNA Ends)", Genome Research, vol. 4, 1994, S40-S58.

Case: 24-1324 Document: 42-1 Page: 58 Filed: 03/18/2024

## US 11,519,035 B2

Page 23

#### (56)References Cited

#### OTHER PUBLICATIONS

Garcia Moreira, V., et al., "Cell-Free DNA as a Noninvasive Acute Rejection Marker in Renal Transplantation", Clinical Chemistry, 2009, 1958-1966

Guo, H, et al., "A Specific and Versatile Genome Walking Technique", Gene, vol. 381, 2006, 18-23.

Hyndman, D L, et aL, "PCR Primer Design", Methods in Molecular Biology, vol. 226, Second Edition, 2003, 81-88.

Illumina, "HiSeq 2500 Sequencing System", System Specification Sheet: Sequencing, 2015, 4 pgs.
Illumina, "TruSeq™ RNA and DNA Library Preparation Kits v2",

Data Sheet: Illumina® Sequencing, 2014, 4.

Iskow, R. C., et al., "Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons", Cell, vol. 141, Jun. 25, 2010, 1253-1261.

Juppner, H., et al., "Functional Properties of the PTH/PTHrP Receptor", Bone, Aug. 14, 1995, 39S-42S.

Kane, M., et al., "Application of Less Primer Method to Commer-

cial Kits", Forensic Science International: Genetics Supplement Series, vol. 1, Issue 1, 2008, 41-43.

Kivioja, T., et al., "Counting absolute number of molecules using unique molecular identifiers", Nature Proceedings, Apr. 14, 2011, 18 pgs

Kivioja, T., et al., "Counting absolute number of molecules using unique molecular identifiers", Nature Methods, Jan. 2012, 72-76. Konfortov, B A, et al., "A High-Resolution Happy Map of Dictyostelium discoideum Chromosome 6", Genome Research, vol. 10, No. 11, Nov. 2000, 1737-1742.

Kopreski, MS, et al., "Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer", British Journal of Cancer, 1997, 1293-1299.

Kuhn, H., et al., "Rolling-circle amplification under topological constraints", Nucleic Acids Research, 2002, 574-580.

Lavrentieva, I, et al., "High Polymorphism Level of Genomic Sequences Flanking Insertion Sites of Human Endogenous Retroviral Long Terminal Repeats", FEBS Letters, vol. 443, No. 3, Jan. 29, 1999, 341-347.

Lee, J , et al., "Anchored Multiplex PCR Enables Sensitive and Specific Detection of Variants in Circulating Tumor DNA by Next-Generation Sequencing", DOI:https://doi.org/10.1016/j.cancergen. 2017.04.049, Cancer Genetics 214-215, 2017, 47.

Metzker, M. L., et al., "Polymerase Chain Reaction", Encyclopedia of Medical Devices and Instrumentation, vol. 5, Second Edition, 2006, 380-387.

Metzker, M. L., et al., "Quantitation of Mixed-Base Populations of HIV-1 Variants by Automated DNA Sequencing with Bodipy" Dye-Labeled Primers, BioTechniques, Sep. 1998, 446-462.

Meyer, M, et al., "Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing", Cold Spring Harbor Protocols, vol. 2010, Issue 6, Jun. 2010, 1-10.
Niamlos, H. M., et al., "Noninvasive Detection of ctDNA Reveals

Intratumor Heterogeneity and Is Associated with Tumor Burden in Gastrointestinal Stromal Tumor", Molecular Cancer Therapeutics, vol. 17, No. 11, 2018, 2473-2480.

Nawroz, H, et al., "Microsatellite Alterations in Serum DNA of Head and Neck Cancer Patients", Nature Medicine, vol. 2, No. 9, Sep. 1996, 1035-1037.

Dhara, O, et al., "One-sided Polymerase Chain Reaction: The Amplification of cDNA", Proceedings of the National Academy of Sciences, vol. 86, 1989, 5673-5677.

Ohira, T., et al., "Tumor volume determines the feasibility of cell-free DNA sequencing for mutation detection in non-small cell lung cancer", Cancer Science, 2016, 1660-1666.

Olivarius, S, et al., "High-throughput Verification of Transcriptional starting Sites by Deep-Race", Bio Techniques, vol. 46, No. 2, Feb. 2009, 130-132.

Perakis, S., et al., "Advances in Circulating Tumor DNA Analysis", Advances in Clinical Chemistry, 2017, 73-153.

Profitt, J, et al., "Isolation and Characterisation of Recombination Events Involving Immunoglobulin Heavy Chain Switch Regions in

Multiple Myeloma Using Long Distance Vectorette PCR (Ldvper)", Leukemia, vol. 13, No. 7, Jul. 1999, 1100-1107.

Rabinowitz, M., "A System and Method for Integrating, Validating and Applying Genetic and Clinical Data to Enhance Medical Decisions", Nov. 29, 2005, 155 pgs.

Santalucia, Jr., J., "Physical Principles and Visual-OMP Software for Optimal PCR Design", Methods in Molecular Biology, vol. 402, 2007, 3-33.

Schoske, R , et al., "Multiplex PCR Design Strategy used for the Simultaneous Amplification of 10 Y Chromosome Short Tandem Repeat (STR) Loci", Analytical and Bioanalytical Chemistry, vol. 375, 2003, 333-343.

Shapero, M. H., et al., "MARA: A Novel Approach for Highly Multiplexed Locus-specific SNP Genotyping Using High-density DNA Oligonucleotide Arrays", Nucleic Acids Research, vol. 32, No. 22, 2004, 1-9.

Tounta, G, et al., "Non-invasive prenatal diagnosis using cell-free fetal nucleic acids in maternal plasma: Progress Overview beyond predictive and personalized diagnosis", EPMA Journal, vol. 2, Issue 2, 2011, 163-171.

Treff, N. R., et al., "Single Cell Whole Genome Amplification Technique Significantly Impacts the Accuracy and Precision of Microarray Based 23 Chromosome Aneuploidy Screening", Poster Presentations Preimplantation Genetic Diagnosis, Sep. 01 2007,

Troutt, et al., "Ligation-anchored PCR: A Simple Amplification Technique with Single-sided Specificity", Proceedings of the National Academy of Sciences, vol. 89, Oct. 1992, 9823-9825

Tufan, N L, et al., "Analysis of Cell-Free Fetal DNA from Maternal Plasma and Serum Using a Conventional Multiplex PCR: Factors Influencing Success", The Turkish Journal of Medical Sciences, vol.

Zheng, Z, et al., "Anchored Multiplex PCR for Targeted Nextgeneration Sequencing", Nature Medicine, vol. 20, No. 12, Dec. 2014, 1479-1486.

Zhong, X Y, et al., "Detection of Fetal Rhesus D and Sex Using Fetal DNA from Maternal Plasma by Multiplex Polymerase Chain Reaction", British Journal of Obstetrics and Gynaecology, vol. 107, Jun. 2000, 766-769.

Abd-Elsalam, Kamel A., "Bioinformatic Tools and Guideline for PCR Primer Design", African Journal of Biotechnology, vol. 2, 2003, pp. 91-95.

Adalsteinsson, V. A. et al., "Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors", Nature Communications, vol. 18, No. 1324, 2017, 13 pages.

Adinolfi, M. et al., "Rapid Detection of Aneuploidies by Microsatellite and the Quantitative Fluorescent Polymerase Chain Reaction", Prenatal Diagnosis, vol. 17, No. 13, 1997, 1299-1311. Arandjelovic, M. et al., "Two-Step Multiplex Polymerase Chain

Reaction improves the Speed and Accuracy of Genotyping Using DNA from Noninvasive and Museum Samples", Molecular Ecology Resources, vol. 9, 2009, pp. 28-36.

Balavoine, Guillaume, "Identification of Members of Several Homeobox Genes in a Planarian Using a Ligation-Mediated Polymerase Chain Reaction Technique", Nucleic Acids Research, vol. 24, 1996, pp. 1547-1553.

Bartlett, John M. et al., "PCR Protocols", PCR Protocols, vol. 226, 2003, 519 pages.

Benjamini, Y. et al., "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing", Journal of the Royal Statistical Society, Series B (Methodological), vol. 57, No. 1, 1995, 289-300.

Brastianos, P. K. et al., "Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets", Cancer Discovery, vol. 5, Sep. 26, 2015, 1164-1177.

Chavali, Sreenivas et al., "Oligonucleotide Properties Determination and Primer Designing: a Critical Examination of Predictions", Bioinformatics, vol. 21, 2005, pp. 3918-3925.

Chen, Bing-Yuan et al., "PCR Cloning Protocols", PCR Cloning

Protocols, vol. 192, 2002, 434 pages. Cunningham, K. S. et al., "An approach to endomyocardial biopsy interpretation", Journal of Clinical Pathology, vol. 59, No. 2, Mar. 2006, 121-129.

## US 11,519,035 B2

Page 24

#### (56) References Cited

#### OTHER PUBLICATIONS

Deb, Mahua et al., "Development of a Multiplexed PCR Detection Method for Barley and Cereal Yellow Dwarf Viruses, Wheat Spindle Streak Virus, Wheat Streak Mosaic Virus and Soil-Borne Wheat Mosaic Virus", Journal of Virological Methods, vol. 148, 2008, pp. 17-24.

Delgado, P. O. et al., "Characterization of cell-free circulating DNA in plasma in patients with prostate cancer", Tumor Biol., vol. 34, 983-986, 2013.

Dorit, D. L., "cDNA Amplification Using One-sided (Anchored) PCR", Current Protocols in Molecular Biology, vol. 17, 1992, pp. 15.6.1-15.6.10.

Dorit, Robert L. et al., "One-sided Anchored Polymerase Chain Reaction for Amplification and Sequencing of Complementary DNA", Methods in Enzymology, vol. 218., 1993, pp. 36-47.

Efron, B. et al., "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy", Statistical Science, vol. 1, No. 1, 1986, 54-77.

Elnifro, Elfath M., "Multiplex PCR: Optimization and Application in Diagnostic Virology", Clinical Microbiology Reviews, vol. 13, 2000, pp. 559-570.

Erijman, Ariel et al., "Transfer-PCR (TPCR): A Highway for DNA Cloning and Protein Engineering", Journal of Structural Biology, vol. 175, 2011, pp. 171-177.

European Commission, "The 7th International Conference on Circulating Nucleic Acids in Plasma and Serum (CNAPS VII) in Madrid—Spain", The International Conference on Circulating Nucleic Acids in Plasma and Serum, Oct. 24, 2011, 2 pgs.

Fan, H. C. et al., "In Principle Method for Noninvasive Determination of the Fetal Genome", Nat. Prec., 2010, 16 pgs.

Gadi, V. K. et al., "Soluble Donor DNA Concentrations in Recipient Serum Correlate with Pancreas-Kidney Rejection", Clinical Chemistry, vol. 52, No. 3, 2006, 379-382

istry, vol. 52, No. 3, 2006, 379-382. Gao, Ming et al., "Characterization of dull1, a Maize Gene Coding for a Novel Starch Synthase", The Plant Cell, vol. 10, 1998, pp. 309-412

Gielis, E. M. et al., "Cell-Free DNA: An Upcoming Biomarker in Transplantation", American Journal of Transplantation, vol. 15, 2015, 2541-2551.

Go, A. T. et al., "Non-invasive aneuploidy detection using free fetal DNA and RNA in maternal plasma: recent progress and future possibilities", Human Reproduction Update, vol. 17, No. 3, 2011, 372-382.

Gwee, Pai-Chung et al., "Simultaneous Genotyping of Seven Singlenucleotide Polymorphisms in the Mdr1 Gene by Single-tube Multiplex Minisequencing", Pai-Chung Gwee. et al., "Simultaneous Genotyping of Seven Single-nucleotide Polymorphisms in the Mdr1 Gene by Single-tube Multiplex Minisequencing", Clinical chemistry, Apr. 2003, vol. 49, Issue. 3, pp. 672-676., Apr. 1, 2003, 672-676. Hahn, S. et al., "Quantification of Circulating DNA: in the Preparation Lies the Rub", Clinical Chemistry, vol. 47, No. 9, 2001, 1577-1578.

Halford, William P., "The Essential Prerequisites for Quantitative RT-PCR", Nature Biotechnology, vol. 17, 1999, 1 page.

Handley, D. et al., "Noninvasive prenatal chromosomal aneuploidy detection using plasma cell-free nucleic acid", Expert Rev Obstet. Gynecol, vol. 5, No. 5, 2010, 581-590. Heaton, Paul R. et al., "Heminested PCR Assay for Detection of Six

Heaton, Paul R. et al., "Heminested PCR Assay for Detection of Six Genotypes of Rabies and Rabies-related Viruses", Journal of Clinical Microbiology, vol. 35, 1997, pp. 2762-2766.

Heidary, M. et al., "The dynamic range of circulating tumor DNA in metastatic breast cancer", Breast Cancer Research, vol. 16, No. 421, 2014, 10 pages.

Henegariu, O. et al., "Multiplex PCR: Critical Parameters and Step-by-Step Protocol", Biotechniques, vol. 23, 1997, pp. 504-511. Hidestrand, M. et al., "Highly Sensitive Noninvasive Cardiac Transplant Rejection Monitoring Using Targeted Quantification of Donor-Specific Cell-Free Deoxyribonucleic Acid", Journal of the American College of Cardiology, vol. 63, No. 12, 1224-1226, 2014. Hosmillo, Myra D. et al., "Development of Universal SYBR Green Real-time RT-PCR for the Rapid Detection and Quantitation of Bovine and Porcine Toroviruses", Journal of Virological Methods, vol. 168, 2010, pp. 212-217.

Hu, Hao et al., "Mutation Screening in 86 Known X-linked Mental Retardation Genes by Droplet-based Multiplex Pcr and Massive Parallel Sequencing", Hao Hu. et al., "Mutation Screening in 86 Known X-linked Mental Retardation Genes by Droplet-based Multiplex Pcr and Massive Parallel Sequencing", Hugo J, Dec. 2009, vol. 3, pp. 41-49., Dec. 1, 2009, 41-49.

Huang, D. J. et al., "Reliable detection of Trisomy 21 using MALDI-TOF mass spectrometry", Genetics in Medicine, vol. 8, Nov. 2006, 728-734.

Jen, J. et al., "An Overview on the Isolation and Analysis of Circulating Tumor DNA in Plasma and Serum", Annals New York Academy of Sciences, 2000, 8-12.

Jung, K. et al., "Cell-free DNA in the blood as a solid tu1nor biomarker—A critical appraisal of the literature", Clinica Chimica Acta, vol. 411, 2010, 1611-1624.

Kalendar, Ruslan et al., "Java Web Tools for PCR, in Silico PCR, and Oligonucleotide Assembly and Analysis", Genomics, vol. 98, 2011, pp. 137-144.

Kane, M., "Application of Less Primer Method to Multiplex PCR", International Congress Series, vol. 1288, 2006, pp. 694-696. Kennedy, S. R. et al., "Detecting ultralow-frequency mutations by

Kennedy, S. R. et al., "Detecting ultralow-frequency mutations by Duplex Sequencing", Nature Protocols, vol. 9, No. 11, 2014, 2586-2606.

Kibbe, Warren A., "Oligocalc: An Online Oligonucleotide Properties Calculator", Nucleic Acids Research, vol. 35, 2007, pp. W43-W46

Koide, K. et al., "Fragmentation of cell-free fetal DNA in plasma and urine of pregnant women", Prenatal Diagnosis, vol. 25, 2005, 604-607.

Koressaar, Triinu et al., "Enhancements and Modifications of Primer Design Program Primer3", Bioinformatics, vol. 23, 2007, pp. 1289-1291

Lardeux, Frederic et al., "Optimization of a Semi-nested Multiplex PCR to Identify Plasmodium Parasites in Wild-Caught Anopheles in Bolivia, and Its Application to Field Epidemiological Studies", Transactions of the Royal Society of Tropical Medicine and Hygiene, vol. 102, 2008, pp. 485-492.

Larsen, J. B. et al., "Single-step Nested Multiplex PCR to Differ-

Larsen, J. B. et al., "Single-step Nested Multiplex PCR to Differentiate Between Various Bivalve Larvae", Marine Biology, vol. 146, 2005, pp. 1119-1129.

2005, pp. 1119-1129. Li, Ying et al., "Ready detection of donor-specific single-nucleotide polymorphisms in the urine of renal transplant recipients by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", Ying Li. et al., "Ready detection of donor-specific single-nucleotide polymorphisms in the urine of renal transplant recipients by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", Clin Chem,Oct. 2005,vol. 51,Issue.10,pp. 1903-04, Oct. 1, 2005, 1903-1904.

Loh, Elwyn , "Anchored PCR: Amplification with Single-sided Specificity", Methods, vol. 2, 1991, pp. 11-19.
Marguiles, M. et al., "Genome Sequencing in Open Microfabricated

Marguiles, M. et al., "Genome Sequencing in Open Microfabricated High Density Picoliter Reactors", Nature, vol. 437, No. 7057, Sep. 15, 2005, 376-380.

Marianes, Alexis E. et al., "Targets of Somatic Hypermutation within Immunoglobulin Light Chain Genes in Zebrafish", Immunology, vol. 132, 2010, pp. 240-255.

Matsubara, T. et al., "Pantropic Retroviral Vectors Integrate and Express in Cells of the Malaria Mosquito, Anopheles Gambiae", PNAS, vol. 93, 1996, pp. 6181-6185.

PNAS, vol. 93, 1996, pp. 6181-6185. Messmer, Trudy O. et al., "Application of a Nested, Multiplex PCR to Psittacosis Outbreaks", Journal of Clinical Microbiology, vol. 35, No. 8, 1997, pp. 2043-2046.

Meuzelaar, Linda S. et al., "Megaplex PCR: A Strategy for Multiplex Amplification", Nature Methods, vol. 4, 2007, pp. 835-837. Miramontes, Pedro et al., "DNA Dimer Correlations Reflect in Vivo Conditions and Discriminate Among Nearest-neighbor Base Pair Free Energy Parameter Measures", Physica A, vol. 321, 2003, pp. 577-586.

## US 11,519,035 B2

Page 25

#### (56) References Cited

#### OTHER PUBLICATIONS

Mitra, S. et al., "Chapter 4 Classification Techniques", Introduction to Machine Learning and Bioinformatics, First Edition, 2008, 101-127

Moreau, Valerie et al., "Zip Nucleic Acids: New High Affinity Oligonucleotides as Potent Primers for PCR and Reverse Transcription", Nucleic Acids Research, vol. 37, No. 19, e130, 2009, 14 pages

Nishiwaki, Morie et al., "Genotyping of Human Papillomaviruses by a Novel One-step Typing Method With Multiplex PCR and Clinical Applications", Journal of Clinical Microbiology, vol. 46, 2008, pp. 1161-1168.

Norton, S. E. et al., "A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR", Clin Biochem., vol. 46, No. 15, Oct. 2013, 1561-1565.

O'Connell, G. C. et al., "High Interspecimen Variability in Nucleic Acid Extraction Efficiency Necessitates the Use of Spike-In Control for Accurate qPCR-based Measurement of Plasma Cell-Free DNA Levels", Lab Medicine, vol. 48, 2017, 332-338. Owczarzy, Richard et al., "Melting Temperatures of Nucleic Acids:

Owczarzy, Richard et al., "Melting Temperatures of Nucleic Acids: Discrepancies in Analysis", Biophysical Chemistry, vol. 117, 2005, pp. 207-215.

Paik, P. K. et al., "Next-Generation Sequencing of Stage IV Squamous Cell Lung Cancers Reveals an Association of P13K Aberrations and Evidence of Clonal Heterogeneity in Patients with Brain Metastases", Cancer Discovery, vol. 5, Apr. 30, 2015, 610-621.

Palka-Santini, Maria et al., "Large Scale Multiplex PCR Improves Pathogen Detection by DNA Microarrays", BMC Microbiology, vol. 9, No. 1, 2009, 14 pages.

Panjkovich, Alejandro et al., "Comparison of Different Melting Temperature Calculation Methods for Short DNA Sequences", Bioinformatics, vol. 21, 2005, pp. 711-722.

Pfaffl, Michael W., "Quantification Strategies in Real-time PCR", A-Z of quantitative PCR, 2004, pp. 87-112.

"QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook", QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook, Feb. 2003 ("Qiagen (2003)"), 2003, 68 pages.

Ross, P. et al., "Quantitative Approach to Single-Nucleotide Polymorphism Analysis Using MALDI-TOF Mass Spectrometry", BioTechniques, vol. 29, Sep. 2000, 620-629.

Rubio, J. M. et al., "Semi-nested, Multiplex Polymerase Chain Reaction for Detection of Human Malaria Parasites and Evidence of Plasmodium Vivax Infection in Equatorial Guinea", The American Journal of Tropical Medicine and Hygiene, vol. 60, 1999, pp. 183-187.

Schaaf, C. P. et al., "Copy Number and SNP Arrays in Clinical Diagnostics", Annu. Rev. Genomics Hum. Genet., vol. 12, 2011, 25-51.

Schwarzenbach, H. et al., "Detection and Characterization of Circulating Microsatellite-DNA in Blood of Patients with Breast Cancer", Ann. N. Y. Acad. Sci., vol. 1022, 2004, 25-32.

Shyamala, Venkatakrishna et al., "Genome Walking by Single-Specific-Primer Polymerase Chain Reaction: SSP-PCR", Gene, vol. 84, 1989, pp. 1-8.

Singh, Vinayak K. et al., "PCR Primer Design", Molecular Biology Today, vol. 2, 2001, pp. 27-32. Sorenson, G. D. et al., "Soluble Normal and Mutated DNA Sequences

Sorenson, G. D. et al., "Soluble Normal and Mutated DNA Sequences from Single-Copy Genes in Human Blood", Cancer Epdemiology,

Biomarkers & Prevention, vol. 3, Jan./Feb. 1994, 67-71. Spindler, K.L. G. et al., "Cell-free DNA in healthy individuals, noncancerous disease and strong prognostic value in colorectal cancer", International Journal of Cancer, vol. 135, 2014, 2984-2991. Spindler, K.-L. G. et al., "Cell-Free DNA in Metastatic Colorectal Cancer: A Systematic Review and Meta-Analysis", The Oncologist, vol. 22, 2017, 1049-1055.

Stewart, S. et al., "Revision of the 1990 Working Formulation for the Standardization of Nomenclature in the Diagnosis of Heart Rejection", The Journal of Heart and Lung Transplantation, vol. 24, No. 11, Nov. 2005, 1710-1720. Stone, J. P. et al., "Ex Vivo Normothermic Perfusion Induces Donor-Derived Leukocyte Mobilization and Removal Prior to Renal Transplantation", Kidney Int Rep., vol. 1, No. 4, Aug. 6, 2016, 230-239.

Swarup, V. et al., "Circulating (cell-free) nucleic acids—A promising, non-invasive tool for early detection of several human diseases", FEBS Letters, vol. 581, 2007, 795-799.

Thompson, J. C. et al., "Detection of Therapeutically Targetable Driver and Resistance Mutations in Lung Cancer Patients by Next-Generation Sequencing of Cell-Free Circulating Tumor DNA", Clin Cancer Res, vol. 22, No. 23, Dec. 1, 2016, 5772-5782. Thornton, Brenda et al., "Real-time PCR (qPCR) Primer Design

Thornton, Brenda et al., "Real-time PCR (qPCR) Primer Design Using Free Online Software", Biochemistry and Molecular Biology Education, vol. 39, 2011, pp. 145-154.

Urbanova, M. et al., "Circulating Nucleic Acids as a New Diagnostic Tool", Cellular & Molecular Biology Letters, vol. 15, 2010, 242-259.

Vallone, Peter M. et al., "Demonstration of Rapid Multiplex PCR Amplification Involving 16 Genetic Loci", Forensic Science International: Genetics, vol. 3, 2008, pp. 42-45.

Vanneste, Marion et al., "Functional Genomic Screening Independent of the Company of th

Vanneste, Marion et al., "Functional Genomic Screening Independently Identifies CUL3 as a Mediator of Vemurafenib Resistance via Src-RAC1 Signaling Axis", Frontiers in Oncology, vol. 10, 2020, 16 pages.

pages. Vlaminck, I. D. et al., "Circulating Cell-Free DNA Enables Non-invasive Diagnosis of Heart Transplant Rejection", Sci Transl Med., vol. 6, No. 241, Jun. 18, 2018, 26 pages.

Von Ahsen, Nicolas et al., "Oligonucleotide Melting Temperatures under PCR Conditions: Nearest-Neighbor Corrections for Mg2+, Deoxynucleotide Triphosphate, and Dimethyl Sulfoxide Concentrations with Comparison to Alternative Empirical Formulas", Clinical Chemistry, vol. 47, 2001, pp. 1956-1961.

Wartell, Roger M. et al., "Thermal Denaturation of DNA Molecules: A Comparison of Theory with Experiment", Physics Reports, vol. 126, 1985, pp. 67-107.

Watt, Heather L., "Sex Diagnosis of Preimplantation Porcine Embryos through PCR Amplification of The SRY Gene", Sex Diagnosis of Preimplantation Porcine Embryos Through PCR Amplification of the Sry Gene (1998) ("Watt (1998)"), 1998, 151 pages. Wei, Ting et al., "Novel Approaches to Mitigate Primer Interaction and Eliminate Inhibitors in Multiplex PCR, Demonstrated Using an Assay for Detection of three Strawberry Viruses", Journal of Virological Methods, vol. 151, 2008, pp. 132-139.

Wilkinson, Sarah T. et al., "Decreased MHC Class II Expression in Diffuse Large B-Cell Lymphoma does not Correlate with CPG Methylation of Ciita Promoters III and IV", Leuk Lymphoma, vol. 50, 2009, pp. 1875-1878.

Witherspoon, David J. et al., "Mobile Element Scanning (Me-scan) by Targeted High-throughput Sequencing", BMC Genomics, vol. 410, 2010, 15 pages.

410, 2010, 15 pages.

Wong, K. H. et al., "Multiplex Illumina Sequencing Using DNA Barcoding", Current Protocols in Molecular Biology, vol. 101, Jan. 2013, 7.11.1-7.11.11.

Yilmaz, A. et al., "Comparative Evaluation of Left and Right Ventricular Endomyocardial Biopsy", Circulation, vol. 122, No. 9, Aug. 31, 2010, 900-909.

Yuanxin, Yan et al., "T-linker-specific Ligation PCR (T-linker Pcr): An Advanced PCR Technique for Chromosome Walking or for Isolation of Tagged DNA Ends", Nucleic Acids Research, vol. 31, No. 12, e68, 2003, 7 pages.

Bau, Stephan et al., "Targeted next-generation sequencing by spe-

Bau, Stephan et al., "Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays", Anal Bioanal Chem, vol. 393, 2009, 171-175.

Goncalves-Primo, a. et al., "Investigation of Apoptosis-Related Gene Expression Levels in Preimplantation Biopsies as Predictors of Delayed Kidney Graft Function", Transplantation, vol. 97, No. 12, Jun. 27, 2014.

Nakamura, N. et al., "Ex Vivo Liver Perfusion with Arterial Blood from a Pig with Ischemic Liver Failure", Artificial Organs, vol. 23, No. 2, 1999, 153-160.

Nui, A. et al., "The Functional Integrity of a Normothermic Perfusion System Using Artificial Blood in Pig Liver", Journal of Surgical Research, Vo. 131, 2006, 189-198.

## US 11,519,035 B2

Page 26

#### (56) References Cited

#### OTHER PUBLICATIONS

Olerup, O. et al., "HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological Dr typing in clinical practice including donor-recipient matching in cadaveric transplantation", Tissue Antigens, vol. 39, No. 5, May 1992, 225-235.

Paruzynski, A. et al., "Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing", Nature Protocols, vol. 5, No. 8, Jul. 8, 2010, 1379-1395.

Alizadeh, Mehdi et al., "Quantitative Assessment of Hematopoietic Chimerism after Bone Marrow Transplantation by Real-time Quantitative Polymerase Chain Reaction", Blood, vol. 99, No. 12, Jun. 15, 2002, 4618-4625.

Ansorge, Wilhelm J., "Next-generation DNA Sequencing Techniques", New Biotechnology, vol. 25, No. 4, Feb. 2, 2009, 195-203. Avent, Neil D. et al., "Cell-free Fetal DNA in the Maternal Serum and Plasma: Current and Evolving Applications", Current Opinion in Obstretrics and Gynecology, vol. 21, No. 2, Apr. 1, 2009, 175-179.

Ayala, et al., "Long-Term Follow-Up of Donor Chimerism Tolerance After Human Liver Transplantation", Liver Transplantation, vol. 15, No. 6, May 28, 2009, 581-591.

Balduini, et al., "Utility of Biochemical Markers in the Follow-up Heart Transplant Recipients", Transplantation Proceedings, vol. 35, No. 8, Dec. 1, 2003, 3075-3078.

Barbazuk, et al., "SNP Discovery via 454 Transcriptome Sequencing", The Plant Journal, vol. 51, Jul. 27, 2007, 910-918.

Baxter-Lowe, et al., "Tracking Microchimeric DNA in Plasma to Diagnose and Manage Organ Transplant Rejection", Clinical Chemistry, vol. 52, No. 4, Apr. 1, 2006, 559-561.

Beck, et al., "Next Generation Sequencing of Serum Circulating Nucleic Acids from Patients with Invasive Ductal Breast Cancer Reveals Differences to Healthy and Nonmalignant Controls", Molecular Cancer Research, vol. 8, No. 3, Mar. 1, 2010, 335-342.

Belostotsky, Dmitry A. et al., "Plant Systems Biology", Methods in Molecular Biology, vol. 553, Aug. 25, 2009, 3-408.

Bender, et al., "A Multiplex SNP Typing Approach for the DNA Pyrosequencing Technology", International Congress Series, vol. 1288, Apr. 20, 2006, 73-75.

Bentley, et al., "High-resolution, High-throughput HLA Genotyping by Next-generation Sequencing", Tissue Antigens, vol. 74, No. 5, Nov. 1, 2009, 393-403.

Bordoni, et al., "Evaluation of Human Gene Variant Detection in Amplicon Pools by the GS-FLX Parallel Pyrosequencer", BMC Genomics, vol. 9, Oct. 8, 2008, 1-8.

Brockman, et al., "Quality Scores and SNP Detection in Sequencingby-synthesis Systems", Genome Research, vol. 18, No. 5, May 1, 2008, 763-770.

Bustamante-Aragones, Ana et al., "New Strategy for the Prenatal Detection/Exclusion of Paternal Cystic Fibrosis Mutations in Maternal Plasma", Journal of Cystic Fibrosis, vol. 7, Issue 6, Nov. 1, 2008, 505,510

Butler, et al., "Cardiovascular Magnetic Resonance in the Diagnosis of Acute Heart Transplant Rejection: a Review", Journal of Cardiovascular Magnetic Resonance, vol. 11, No. 1, Mar. 12, 2009, 1-11.

Chan, Allen K. et al., "Cell-free Nucleic Acids in Plasma, Serum and Urine: A New Tool in Molecular Diagnosis", Annals of Clinical Biochemistry, vol. 40, Issue 2, Mar. 1, 2003, 122-130.

Church, et al., "Multiplex DNA Sequencing", Science, vol. 240, No. 4849, Apr. 8, 1988, 185-188.

Crespo-Leiro, et al., "Gene Expression Profiling for Monitoring Graft Rejection in Heart Transplant Recipients", Transplantation Proceedings, vol. 41, No. 6, Jul. 1, 2009, 2240-2243.

Dahl, et al., "Multigene Amplification and Massively Parallel Sequencing for Cancer Mutation Discovery", Proceedings of the National Academy of Sciences, vol. 104, No. 22, May 29, 2007, 9387-9392.

Dambrin, et al., "A New Rejection Criteria in the Heterotopically Placed Rat Heart by Non-invasive Measurement of Dp/Dtmax", The Journal of Heart and Lung Transplantation, vol. 18, No. 6, Jun. 18, 1999, 524-531.

Doostzadeh, et al., "High Throughput Automated Allele Frequency Estimation by Pyrosequencing", PLoS One, vol. 3, No. 7, Jul. 16, 2008. 1-4.

Fitzgerald, "Intravascular Ultrasound Imaging of Coronary Arteries: Is Three Layers the Norm?", Circulation, vol. 86, No. 1, Jul. 1, 1992, 154-158.

Fournie, et al., "Plasma DNA As a Marker of Cancerous Cell Death. Investigations in Patients Suffering From Lung Cancer and in Nude Mice Bearing Human Tumours", Cancer Letters, vol. 91, No. 2, May 8, 1995, 221-227.

Fredriksson, M et al., "Assessing Hematopoietic Chimerism After Allogeneic Stem Cell Transplantation by Multiplexed SNP Genotyping Using Microarrays and Quantitive Analysis of SNP Alleles", Leukemia, vol. 18, Issue 2, Dec. 4, 2003, 255-266.

Fu, Yao-Wen et al., "Presence of Donor-and-recipientderived Dna Microchimerism in the Cell-free Blood Samples of Renal Transplantation Recipients Associates With the Acceptance of Transplanted Kidneys", Asian Journal of Andrology, vol. 8, No. 4, Jul. 1, 2006, 477-482.

Gao, et al., "Relation of Donor Age and Preexisting Coronary Artery Disease on Angiography and Intracoronary Ultrasound to Later Development of Accelerated Allograft Coronary Artery Disease", The American Journal of Cardiology, vol. 29, No. 3, Mar. 1, 1997, 623-629.

Geifman-Holtzman, et al., "Prenatal Diagnosis: Update on Invasive Versus Noninvasive Fetal Diagnostic Testing From Maternal Blood", Expert Review of Molecular Diagnostics, vol. 8, No. 6, Nov. 1, 2008, 727-751.

Gineikiene, Egle et al., "Single Nucleotide Polymorphism-based System Improves the Applicability of Quantitative PCR for Chimerism Monitoring", Journal of Molecular Diagnostics, vol. 11, No. 1, Jan. 1, 2009, 66-74.

Gingeras, et al., "Fifty Years of Molecular (DNA/RNA) Diagnostics", Clinical Chemistry, vol. 51, No. 3, Jan. 13, 2005, 661-671. Girnita, Diana M. et al., "Disparate Distribution of 16 Candidate Single Nucleotide Polymorphisms Among Racial and Ethnic Groups of Pediatric Heart Transplant Patients", Transplantation, vol. 82, No. 12, Dec. 27, 2006, 1774-1780.

Gordon, et al., "Disease-Specific Motifs Can Be Identified in Circulating Nucleic Acids From Live Elk and Cattle Infected With Transmissible Spongiform Encephalopathies", Nucleic Acids Research, vol. 37. No. 2, Feb. 1, 2009, 550-556.

Gorringe, et al., "Large-scale Genomic Analysis of Ovarian Carcinomas", Molecular oncology, vol. 3, No. 2, Apr. 1, 2009, 157-164. Gouya, et al., "Coronary Artery Stenosis in High-risk Patients: 64-section Ct and Coronary Angiography-Prospective Study and Analysis of Discordance", Radiology, vol. 252, No. 2, Aug. 1, 2009, 377-385.

Gregory, et al., "Comparison of Sixty-Four-Slice Multidetector Computed Tomographic Coronary Sngiography to Coronary Angiography With Intravascular Ultrasound for the Detection of Transplant Vasculopathy", The American Journal of Cardiology, vol. 98, No. 7, Aug. 4, 2006, 877-884.

Hahn, et al., "Non-invasive Prenatal Diagnostics Using Next Gen-

Hahn, et al., "Non-invasive Prenatal Diagnostics Using Next Generation Sequencing: Technical, Legal and Social Challenges", Expert Opinion on Medical Diagnostics, vol. 6, No. 6, Nov. 1, 2012, 517-528

Hao, T. B. et al., "Circulating cell-free DNA in serum as a biomarker for diagnosis and prognostic prediction of colorectal cancer", British Journal of Cancer, vol. 111, Aug. 26, 2014, 1482-1489. Hoberman, Rose et al., "A Probabilistic Approach for SNP Discov-

Hoberman, Rose et al., "A Probabilistic Approach for SNP Discovery in High-throughput Human Resequencing Data", Genome Research, vol. 19, Jul. 15, 2009, 1542-1552.

Hochberg, et al., "A Novel Rapid Single Nucleotide Polymorphism (SNP)—Based Method for Assessment of Hematopoietic Chimerism After Allogeneic Stem Cell Transplantation", Blood, vol. 101, No. 1, Jan. 1, 2003, 363-369.

## US 11,519,035 B2

Page 27

### (56) References Cited

#### OTHER PUBLICATIONS

Hodges, et al., "Genome-wide in Situ Exon Capture for Selective Resequencing", Nature Genetics, vol. 39, No. 12, Nov. 4, 2007, 15221527

Hoffmann, Steven et al., "Donor Genomics Influence Graft Events: The Effect of Donor Polymorphisms on Acute Rejection and Chronic Allograft Nephropathy", Kidney International, vol. 66, No. 4, Oct. 1, 2004, 1686-1693.

Holt, et al., "Detecting SNPS and Estimating Allele Frequencies in Clonal Bacterial Populations by Sequencing Pooled DNA", Bioinformatics, vol. 25, No. 16, Aug. 15, 2009, 2074-2075.

Horai, et al., "Novel Implantable Device to Detect Cardiac Allograft Rejection", Circulation, vol. 120, No. Suppl 1, Sep. 15, 2009, 185-190.

Hubacek, et al., "Detection of Donor DNA After Heart Transplantation: How Far Could It Be Affected by Blood Transfusion and Donor Chimerism?", Transplantation Proceedings, vol. 39, Jun. 1, 2007, 1593-1595.

Illumina, "Genomic Sequencing", Data Sheet: Sequencing, 2010, 38939-38944.

Ingman, et al., "SNP Frequency Estimation Using Massively Parallel Sequencing of Pooled DNA", European Journal of Human Genetics, vol. 17, No. 3, Oct. 15, 2008, 383-386.

Interewicz, B. et al., "DNA Released from Ischemic and Rejecting Organs as an Indicator of Graft Cellular Damage", Annals of Transplantation, vol. 9, No. 2, May 1, 2004, 42-45.

Kapadia, Samir R. et al., "Impact of Intravascular Ultrasound in Understanding Transplant Coronary Artery Disease", Current Opinion in Cardiology, vol. 14, No. 2, Mar. 1, 1999, 1-19.

Karger, et al., "DNA Sequencing by Capillary Electrophoresis", Electrophoresis, vol. 30, Supplement 1, Jun. 1, 2009, 1-11.

Karoui, Noureddine E. et al., "Getting More from Digital SNP Data", Statistics in Medicine, vol. 25, Issue 18, Jan. 5, 2006, 3124-3133.

Kass, et al., "Diagnosis of Graft Coronary Artery Disease", Current Opinion in Cardiology, vol. 22, No. 2, Mar. 1, 2007, 139-145. Kathiresan, Sekar et al., "Genome-wide Association of Early-onset

Myocardial Infarction With Common Single Nucleotide Polymorphisms, Common Copy Number Variants, and Rare Copy Number Variants", Nature Genetics, vol. 41, No. 3, Mar. 1, 2009, 1-23.

Kircher, Martin et al., "Improved Base Calling for the Illumina Genome Analyzer Using Machine Learning Strategies", Genome Biology, vol. 10, Issue 8, Article No. R83, Aug. 14, 2009, 83.2-83.9. Kobashigawa, et al., "Multicenter Intravascular Ultrasound Validation Study Among Heart Transplant Recipients", Journal of the American College of Cardiology, vol. 45, No. 9, May 3, 2005, 1532-1537.

Koboldt, et al., "VarScan: Variant Detection in Massively Parallel Sequencing of Individual and Pooled Samples", Bioinformatics, vol. 25, No. 17, Jun. 19, 2009, 2283-2285.

Koelman, et al., "Donor-derived Soluble HLA Plasma Levels Can

Koelman, et al., "Donor-derived Soluble HLA Plasma Levels Can Not Be Used to Monitor Graft Rejection in Heart Transplant Recipients", Transplant Immunology, vol. 8, No. 1, Mar. 1, 2000, 57-64

Koldehoff, Michael et al., "Quantitative analysis of chimerism after allogeneic stem cell transplantation by real-time polymerase chain reaction with single nucleotide polymorphisms, standard tandem repeats, and Y-chromosome-specific sequences", American Journal of Hematology, vol. 81, No. 10, Jul. 12, 2006, 735-746.

Korn, et al., "Integrated Genotype Calling and Association Analysis of SNPS, Common Copy Number Polymorphisms and Rare CNVS", Nature Genetics, vol. 40, No. 10, Oct. 1, 2008, 1253-1260.

Lambert, et al., "Quantification of Maternal Microchimerism by HLA-Specific Real-time Polymerase Chain Reaction", Arthritis and Rheumatism, vol. 50, No. 3, Mar. 1, 2004, 906-914.

Lavebrat, et al., "Single Nucleotide Polymorphism (SNP) Allele Frequency Estimation in DNA Pools Using Pyrosequencing", Nature Protocols, vol. 1, No. 6, Jan. 11, 2007, 2573-2582.

Lavebratt, Catharina et al., "Pyrosequencing-based SNP Allele Frequency Estimation in DNA Pools", Human Mutation, vol. 23, Issue 1, Dec. 19, 2003, 92-97.

Leamon, John H. et al., "A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions", Electrophoresis, vol. 24, No. 21, Nov. 1, 2003, 3769-3777.

Levsky, Jeffrey M. et al., "Efficacy of Coronary Ct Angiography: Where We Are, Where We Are Going and Where We Want to Be", Journal of Cardiovascular Computed Tomography, vol. 3, Supplement 2, Nov. 2, 2009, s99-s108.

Li, et al., "Detection of SNPs in the Plasma of Pregnant Women and in the Urine of Kidney Transplant Recipients by Mass Spectrometry", Annals of the New York Academy of Sciences, vol. 1075, Sep. 5, 2006, 144-147.

Li, et al., "Mapping Short DNA Sequencing Reads and Calling Variants Using Mapping Quality Scores", Genome Research, vol. 18, No. 11,, Aug. 19, 2008, 1851-1858.

Li, et al., "Multiplex Padlock Targeted Sequencing Reveals Human Hypermutable CpG Variations", Genome Research, vol. 19, No. 9, Jun. 12, 2009, 1606-1615.

Li, et al., "SOAP2: An Improved Ultrafast Tool for Short Read Alignment", Bioinformatics, vol. 25, No. 15, Aug. 1, 2009, 1966-1967.

Li, Ying et al., "Detection of Donor-specific DNA Polymorphisms in the Urine of Renal Transplant Recipients", Clinical Chemistry, vol. 49, No. 4, Apr. 1, 2003, 655-658.

Liljedahl, Ulrika et al., "Detecting Imbalanced Expression of SNP Alleles by Minisequencing on Microarrays", BMC Biotechnology, vol. 4, Article No. 24, Oct. 22, 2004, 1-10.

Lo, et al., "Next-generation Sequencing of Plasma/Serum DNA: an Emerging Research and Molecular Diagnostic Tool", Clinical Chemistry, vol. 55, No. 4, Apr. 1, 2009, 607-608.

Lo, et al., "Presence of Donor-specific DNA in Plasma of Kidney and Liver-transplant Recipients", Lancet, vol. 351, No. 9112, May 2, 1998, 1329-1330.

Lo, Y M. et al., "Circulating Nucleic Acids in Plasma and Serum: An Overview", Annals of the New York Academy of Sciences, vol. 945, Sep. 1, 2001, 1-7.

Lui, Yanni Y. et al., "Circulating DNA in Plasma and Serum: Biology, Preanalytical Issues and Diagnostic Applications", Clinical Chemistry and Laboratory Medicine, vol. 40, No. 10, Oct. 29, 2002, 962-968

Lui, Yanni Y. et al., "Origin of Plasma Cell-Free DNA after Solid Organ Transplantation", Clinical Chemistry, vol. 49, No. 3, Mar. 1, 2003, 495-496.

Lun, Fiona M. et al., "Microfluidics Digital PCR Reveals a Higher Than Expected Fraction of Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 54, No. 10, Aug. 14, 2008, 1664-1672.

Maron, Jill L. et al., "Cell-free Fetal DNA Plasma Extraction and Realtime Polymerase Chain Reaction Quantification", Methods in Molecular Medicine, vol. 132, Aug. 1, 2007, 51-63. Martinez- Lopez, J. et al., "Real-time PCR Quantification of

Martinez- Lopez, J. et al., "Real-time PCR Quantification of Haematopoietic Chimerism after Transplantation: A Comparison Between TaqMan and Hybridization Probes Technologies", International Journal of Laboratory Hematology, vol. 32, Issue 1, Part 1, May 12, 2009, e17-e25.

Martins, et al., "Quantification of Donor-derived DNA in Serum: A New Approach of Acute Rejection Diagnosis in a Rat Kidney Transplantation Model", Transplantation Proceedings, vol. 37, No. 1, Jan. 1, 2005, 87-88.

Milani, et al., "Genotyping Single Nucleotide Polymorphisms by Multiplex Minisequencing Using Tag-arrays", DNA Microarrays for Biomedical Research, vol. 529, Jan. 16, 2009, 215-229.

Moreira, et al., "Increase in and Clearance of Cell-free Plasma DNA In Hemodialysis Quantified By Real-time PCR", Clinical Chemistry and Laboratory Medicine, vol. 44, No. 12, Dec. 13, 2006, 1410-1415.

Neve, B. et al., "Rapid SNP Allele Frequency Determination in Genomic DNA Pools by Pyrosequencing", BioTechniques, vol. 32, No. 5, May 1, 2002, 1138-1142.

Case: 24-1324 Document: 42-1 Page: 63 Filed: 03/18/2024

## US 11,519,035 B2

Page 28

#### References Cited (56)

#### OTHER PUBLICATIONS

Ng, et al., "Multiplex Sequencing of Paired-end Ditags (MS-PET): A Strategy for the Ultra-high-throughput Analysis of Transcriptomes and Genomes", Nucleic Acids Research, vol. 34, No. 12, Jul. 13, 2006, 1-10.

Oeth, et al., "iPLEXTM Assay: Increased Plexing Efficiency and Flexibility for MassARRAY System Through Single Base Primer Extension with Mass-Modified Terminators", Sequenom Applica-

tion Note Doc. No. 8876-006, Apr. 28, 2005, 1-12. Okou, et al., "Microarray-based Genomic Selection for Highthroughput Resequencing", Nature Methods, vol. 4, No. 11, Oct. 14, 2007, 907-909.

Okou, David T. et al., "Combining Microarray-based Genomic Selection (MGS) with the Illumina Genome Analyzer Platform to Sequence Diploid Target Regions", Annals of Human Genetics, vol.

Sequence Diploid Target Regions", Annals of Human Genetics, vol. 73, No. 5, Aug. 6, 2009, 502-513.

Oliver, Dwight H. et al., "Use of Single Nucleotide Polymorphisms (SNP) and Real-time Polymerase Chain Reaction for Bone Marrow Engraftment Analysis", The Journal of Molecular Diagnostics, vol. 2, No. 4, Nov. 1, 2000, 202-208.

Olivier, et al., "The Invader Assay for SNP Genotyping", Mutation Research, vol. 573, No. 1-2, Jun. 3, 2005, 103-110.

Orsouw et al. "Complexity Reduction of Polymorphic Sequences

Orsouw, et al., "Complexity Reduction of Polymorphic Sequences (Crops): A Novel Approach for Large-scale Polymorphism Discovery in Complex Genomes", PLoS ONE, vol. 11:e1172, Nov. 14, 2017, 1-10.

Pakstis, et al., "Candidate SNPs for a Universal Individual Identification Panel", Human Genetics, vol. 121, No. 3-4,, Feb. 27, 2007,

Pakstis, et al., "SNPS for Individual Identification", Forensic Science International, vol. 1, May 22, 2008, 479-481.

Pinard, et al., "Assessment of Whole Genome Amplificationinduced Bias Through High-throughput, Massively Parallel Whole Genome Sequencing", BMC Genomics, vol. 7:216, Aug. 23, 2006,

Pourmand, et al., "Multiplex Pyrosequencing", Nucleic Acid Research, vol. 30, No. 7, Apr. 1, 2002, 1-5. Prabhu, et al., "Overlapping Pools for High-throughput Targeted

Resequencing", Genome Research, vol. 19, May 15, 2009, 1254-

Puszyk, William M. et al., "Noninvasive Prenatal Diagnosis of Aneuploidy Using Cell-free Nucleic Acids in Maternal Blood: Promises and Unanswered Questions", Prenatal Diagnosis, vol. 28, No. 1, Nov. 16, 2007, 1-6.

Raindance Technologies, et al., "RainDance Technologies Introduces the RDT 1000", RainDance Technologies, Nov. 12, 2008.

Ravipati, Goutham et al., "Comparison of Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of Stress Testing Versus 64-Multislice Coronary Computed Tomography Angiography in Predicting Obstructive Coronary Artery Disease Diagnosed by Coronary Angiogr", The American Journal of Cardiology, Coronary Artery Disease. vol. 101, Issue 6, Mar. 15, 2008, 774-775

Roche Diagnostics, et al., "Versatile Nucleic Acid Purification", MagnaPure Manual, Feb. 3, 2012.

Rothberg, et al., "The Development and Impact of 454 Sequencing", Nature Biotechnology, vol. 26, No. 10, Oct. 9, 2008, 1117-1124. Ruschendorf, et al., "Alohomora: A Tool for Linkage Analysis Using 10K SNP Array Data", Bioinformatics Applications Notes, vol. 21, No. 9, Jan. 12, 2005, 2123-2125

Sanger, et al., "Nucleotide Sequence of Bacteriophage Lambda DNA", Journal of Molecular Biology, vol. 162, No. 4, Dec. 25,

Sharples, et al., "Diagnostic Accuracy of Coronary Angiography and Risk Factors for Post-heart-transplant Cardiac Allograft Vasculopathy", Transplantation, vol. 76, No. 4, Aug. 27, 2003,

Smith, et al., "Rapid Whole-genome Mutational Profiling using Next-generation Sequencing Technologies", Genome Research, vol. 18, Sep. 4, 2008, 1638-1642.

Smith, James F. et al., "Cell-free Fetal DNA in Maternal Plasma", Neo Reviews, vol. 9, No. 8, Aug. 1, 2008, e332-e337.

"Application Note: DNA Sequencing", 2006, 1-2.

Spes, et al., "Diagnostic and Prognostic Value of Serial Dobutamine Stress Echocardiography for Noninvasive Assessment of Cardiac Allograft Vasculopathy: A Comparison With Coronary Angiography and Intravascular Ultrasound", Circulation, vol. 100, No. 5, Aug. 3, 1999, 509-515.

Stiller, et al., "Direct Multiplex Sequencing (DMPS)-A Novel Method for Targeted High-thoroughput Sequencing of Ancient and Highly Degraded DNA", Genome Research, vol. 19, No. 10, Jul. 27, 2009, 1843-1848.

Stolerman, Elliot S. et al., "Haplotype structure of the ENPP1 Gene and Nominal Association of the K121Q missense single nucleotide polymorphism with glycemic traits in the Framingham Heart Study", Diabetes, vol. 57, Issue 7, Jul. 1, 2008, 1971-1977.

Takala, et al., "A High-throughput Method for Quantifying Alleles and Haplotypes of The Malaria Vaccine Candidate Plasmodium Falciparum Merozoite Surface Protein-1 19 kDa", Malaria Journal, vol. 5:31, Apr. 20, 2006, 1-10.

Tong, et al., "Diagnostic Developments Involving Cell-free (Circulating) Nucleic Acids", Clinica Chimica Acta, vol. 363, No. (1-2), Aug. 26, 2005, 187-196.

Toshikazu, et al., "Estimation of Haplotype Frequencies, Linkagedisequilibrium Measures, and Combination of Haplotype Copies in Each Pool by Use of Pooled DNA Data", American Journal of Human Genetics, vol. 72, Jan. 17, 2003, 384-398.

Troeger, C. et al., "Approximately Half of the Erythroblasts in Maternal Blood are of Fetal Origin", Molecular Human Reproduction, vol. 5, No. 12, Dec. 1, 1999, 1162-1165.

Tsang, Jason C. et al., "Circulating Nucleic Acids in Plasma/ Serum", Pathology, vol. 39, No. 2, Apr. 1, 2007, 197-207.

Tuzcu, et al., "Intravascular Ultrasound Evidence of Angiographically Silent Progression in Coronary Atherosclerosis Predicts Longterm Morbidity and Mortality After Cardiac Transplantation", The American Journal of Cardiology, vol. 45, No. 9, May 3, 2005, 1538-1542

Umetani, N. et al., "Increased Integrity of Free Circulating DNA in Sera of Patients with Colorectal or Periampullary Cancer: Direct Quantitative PCR for ALU Repeats", Clinical Chemistry, vol. 52, No. 6, 2006, 1062-1069.

Verlaan, et al., "Allele-specific Chromatin Remodeling in the ZPBP22/ GSDMB/ORMDL3 Locus Associated with the Risk of Asthma and Autoimmune Disease", The American Journal of Human Genetics, vol. 85, No. 3, Sep. 11, 2009, 377-393. Verlaan, et al., "Targeted Screening of Cis-Regulatory Variation in

Human Haplotypes", Genome Research, vol. 19, No. 1, Jan. 1, 2009, 118-127.

Voelkerding, et al., "Next-generation Sequencing: From Basic Research to Diagnostics", Clinical Chemistry, vol. 55, No. 4, Apr. 1, 2009, 641-658.

Wasson, Jon et al., "Assessing Allele Frequencies of Single Nucleotide Polymorphisms in DNA Pools by Pyrosequencing Technology", BioTechniques, vol. 32, No. 5, May 1, 2002, 1144-1152.

Wellnhofer, et al., "Angiographic Assessment of Cardiac Allograft Vasculopathy: Results of a Consensus Conference of the Task Force for Thoracic Organ Transplantation of the German Cardiac Society", Transplant International, vol. 23, No. 11, Aug. 19, 2010, 1094-1104

Wiedmann, Ralph T. et al., "SNP Discovery in Swine by Reduced Representation and High Throughput Pyrosequencing", BMC Genetics, vol. 9, Article No. 81, Dec. 4, 2008, 1-7

Wilkening, Stefan et al., "Determination of Allele Frequency in Pooled DNA: Comparison of Three PCR-based Methods", Bio Techniques, vol. 39, No. 6, May 30, 2005, 853-857

Wright, Caroline et al., "Cell-free Fetal Nucleic Acids for Noninvasive Prenatal Diagnosis", PHG Foundation, Jan. 1, 2009, 1-64. Xia, et al., "Simultaneous Quantitative Assessment of Circulating Cell-free Mitochondrial and Nuclear DNA by Multiplex Real-time PCR", Genetics and Molecular Biology, vol. 32, No. 1, Mar. 1, 2009, 20-24.

Case: 24-1324 Document: 42-1 Page: 64 Filed: 03/18/2024

## US 11,519,035 B2

Page 29

#### References Cited (56)

#### OTHER PUBLICATIONS

Xian, et al., "Advances on Circulating Fetal DNA in Maternal Plasma", Chinese Medical Journal, vol. 120, No. 14, Jul. 2, 2007,

Xie, et al., "CNV-SEQ, A New Method to Detect Copy Number Variation Using Highthroughput Sequencing", BMC Bioinformatics, vol. 10:80, Mar. 6, 2009, 1-9.

Xue, et al., "Optimizing the Yield and Utility of Circulating Cellfree DNA From Plasma and Serum", Clinica Chimica Acta, vol. 404, No. 2, Jun. 27, 2009, 100-104.

Yang, Lin et al., "64-MDCT Coronary Angiography of Patients With Atrial Fibrillation: Influence of Heart Rate on Image Quality and Efficacy in Evalution of Coronary Artery Disease", AJR, vol. 193, No. 3, Sep. 1, 2009, 795-801.

Yijen, et al., "Noninvasive Evaluation of Cardiac Allograft Rejection by Cellular and Functional Cardiac Magnetic Resonance", JACC: Cardiovacular Imaging, vol. 2, No. 6, Jun. 1, 2009, 731-741. Zhang, et al., "Diagnosis of Acute Rejection by Analysis of Urinary DNA of Donor Origin in Renal Transplant Recipients", Transplantation Proceedings, vol. 33, No. 1-2, Feb. 2001, 380-381

Zhang, et al., "Use of PCR and PCR-SSP for Detection of Urinary Donor-Origin DNA in Renal Transplant Recipients With Acute Rejection", Chinese Medical Journal, vol. 116, No. 2, Feb. 2003,

Zhang, Kun et al., "Digital RNA Alleotyping Reveals Tissuespecific and Allele-specific Gene Expression in Human", Nature Methods, vol. 6, No. 8, Jul. 20, 2009, 613-618. Zhao, et al., "Urinary Thromboxane B2 in Cardiac Transplant

Patients as a Screening Method of Rejection", Prostaglandins, vol. 54, No. 6, Dec. 1, 1997, 881-889.

Zhong, Xiao Y. et al., "Cell-free DNA in Urine: a Marker for Kidney Graft Rejection, but Not for Prenatal Diagnosis ?", Annals of the New York Academy of Sciences, vol. 945, Sep. 1, 2001, 250-257. Zhou, et al., "Pyrosequencing, A High-throughput Method for Detecting Single Nucleotide Polymorphisms in the Dihydrofolate Reductase and Dihydropteroate Synthetase Genes of Plasmodiym Falciparum", Journal of Clinical Microbiology, vol. 44, No. 11, Nov. 1, 2006, 3900-3910.

Zimmer, et al., "Transplant Coronary Artery Disease", JACC: Cardiovascular Interventions, vol. 3, No. 4, Apr. 1, 2010, 367-377. Lanman, et al., "Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA", Plos One, 2015, 1-27.

Lee, et al., "ERBB2 kinase domain mutation in the lung squamous cell carcinoma", Cancer Letters, 2006, 89-94.

Park, et al., "First-Line Erlotinib Therapy Until and Beyond Response Evaluation Criteria in Solid Tumors Progression in Asian Patients With Epidermal Growth Factor Receptor Mutation-Positive Non-Small-Cell Lung Cancer", JAMA Oncol., 2015, 305-312.

Tseng, Jeng-Sen, et al., "Dynamic Plasma EGFR Mutation Status as a Predictor of EGFR-TKI Efficacy in Patients with fGFR-Mutant Lung Adenocarcinoma", Thorac Oncol., 2015, 603-610. Costa, J.-M. et al., "Fetal RHD genotyping in maternal serum during the first trimester of pregnancy", British Journal of Haematology,

vol. 119, 2002, 255-260

Daniels, G. et al., "Fetal blood group genotyping from DNA from maternal plasma: an important advance in the management and prevention of haemolytic disease of the fetus and newborn", Vox Sanguinis, vol. 87, 2004, 223-232.

Illumina, "HumanOmnil-Quad BeadChip", Illumina DNA Analysis,

Pub. No. 370-21009-007, 2009, 1 page. Illumina, "HumanOmni2.5-8 BeadChips: Next-Generation GWAS Content for Genotyping and CNV Analysis", Data Sheet: DNA Analysis, Pub. No. 370-2011-008, 2011, 1 page. Keshavjee, S. H. et al., "The role of dextran 40 and potassium in

extended hypothermic lung preservation for transplantation", The Journal of Thoracic and Cardiovascular Surgery, vol. 103, No. 2, 1992

Raemdonck, Dirk Van et al., "Ex-vivo lung perfusion", Transplant International, vol. 28, Issue 6, Special Issue: Focus Issue: Machine Perfusion, 2014, 643-656.

Selzner, Markus et al., "Normothermic Ex Vivo Liver Perfusion Using Steen Solution as Perfusate for Human Liver Transplantation: First North American Results", Liver Transplantation, vol. 22, Issue

Tungwiwat, et al., "Non-invasive fetal sex determination using a conventional nested PCR analysis of fetal DNA in maternal plasma", Clinica Chimica Acta, vol. 334, No. 1-2, 2003, 173-177

Ye, et al., "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction", BMC Bioinformatics, 13:134, 2012,

Birkenkamp-Demtroder, et al., "Longitudinal assessment of multiplex patient-specific ctDNA biomarkers in bladder cancer for diagnosis, surveillance and recurrence", Annals of Oncology, Oxford University Press NLD, vol. 29, No. Supplement 8, 2018, viii26. Ehlayel, A. et al., "Emerging monitoring technologies in kidney transplantation", Pediatric Nephrology, vol. 36, 2021, 3077-3087. Glaab, W. E. et al., "A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan) and mismatch amplification mutation assay", Mutation Research, vol. 430, 1999, 12 pgs.

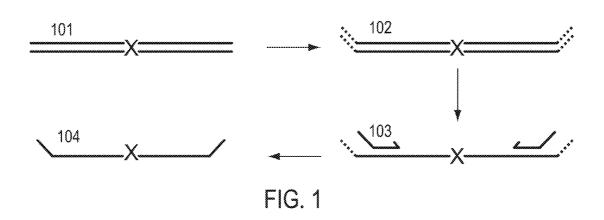
Schutz, E. et al., "Graft-derived cell-free DNA, A noninvasive early rejection and graft damage marker in liver transplantation: a prospective, observational, multicenter cohort study", PLOS Medicine, vol. 14, No. 4, Apr. 25, 2017, 19 pgs. Wangkumhang, P. et al., "WASP: a Web-based Allele-Specific PCR

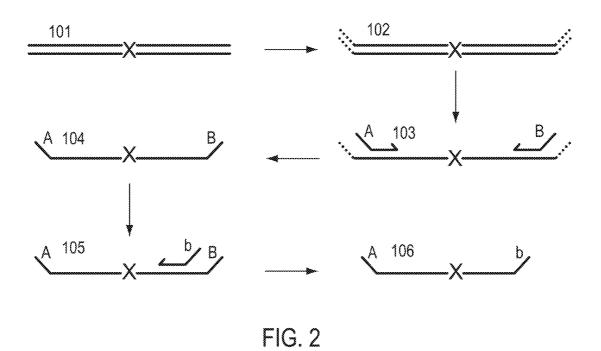
assay designing tool for detecting SNPs and mutations", BMC Genomics, vol. 8, No. 275, Aug. 14, 2007, 9 pgs.

Whitlam, J. B. et al., "Diagnostic application of kidney allograftderived absolute cell-free DNA levels during transplant dysfunction", Am J. Transplant, vol. 19, 2019, 1037-1049.

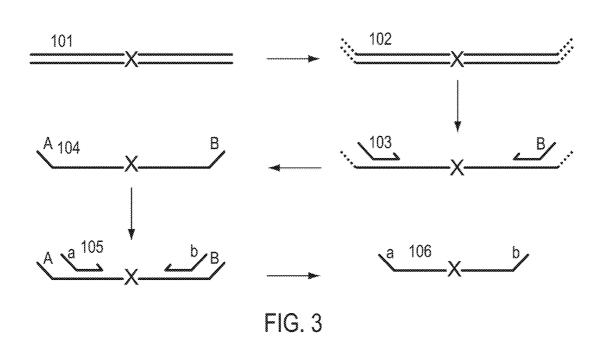
\* cited by examiner

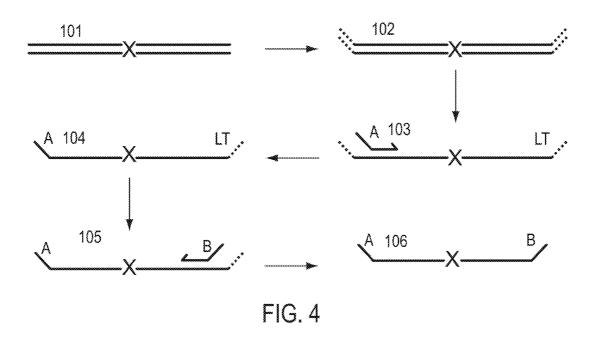
U.S. Patent Dec. 6, 2022 Sheet 1 of 58 US 11,519,035 B2



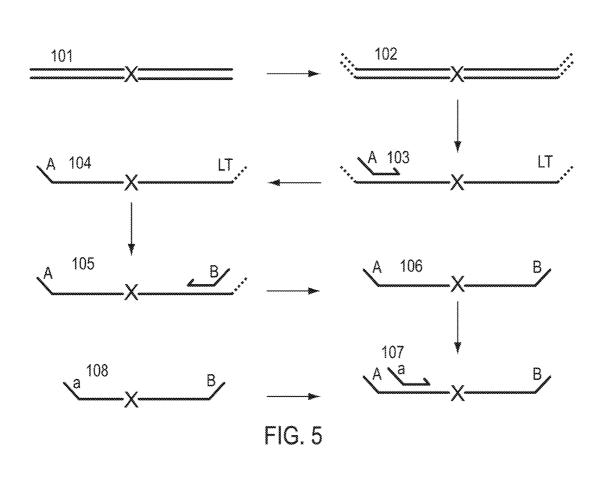


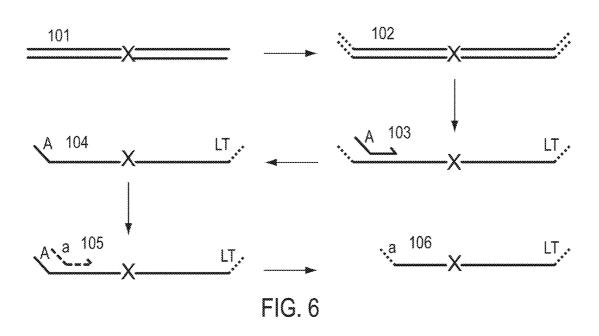
U.S. Patent Dec. 6, 2022 Sheet 2 of 58 US 11,519,035 B2



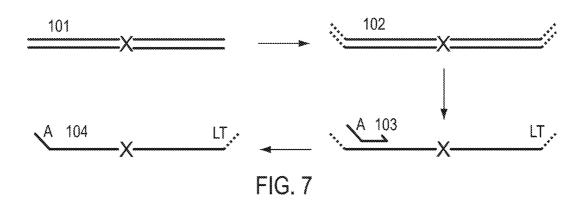


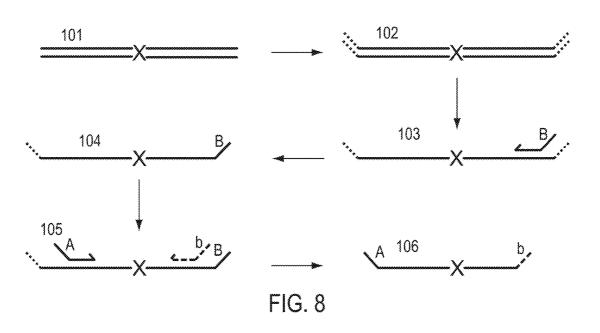
U.S. Patent Dec. 6, 2022 Sheet 3 of 58 US 11,519,035 B2



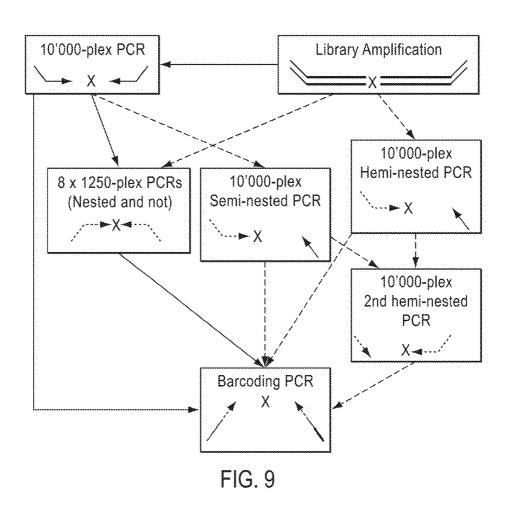


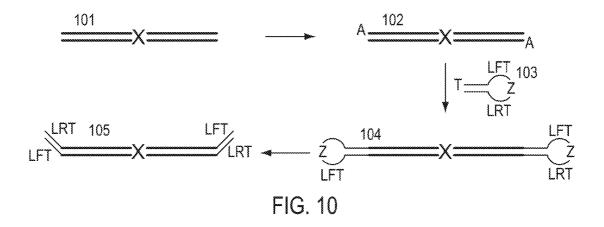
U.S. Patent Dec. 6, 2022 Sheet 4 of 58 US 11,519,035 B2





U.S. Patent Dec. 6, 2022 Sheet 5 of 58 US 11,519,035 B2





U.S. Patent Dec. 6, 2022 Sheet 6 of 58 US 11,519,035 B2

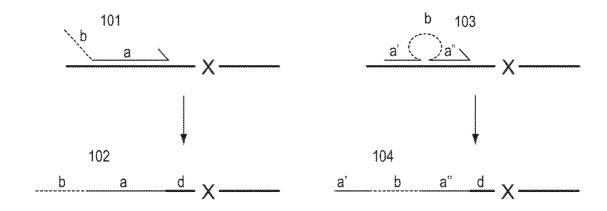
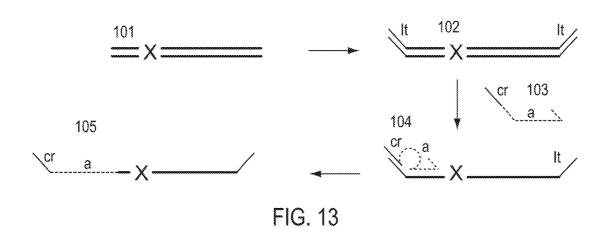
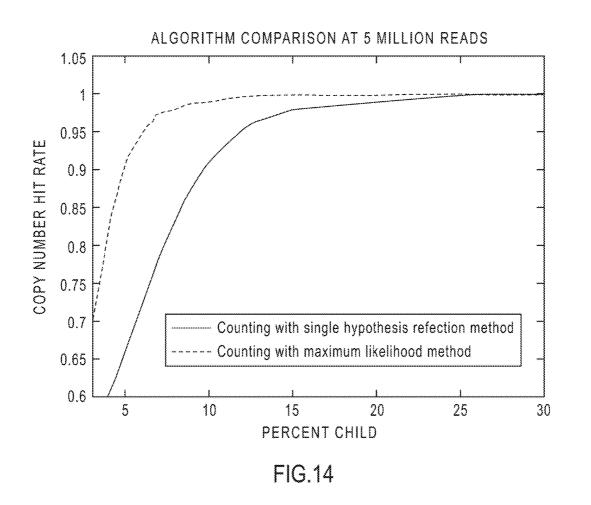


FIG. 11

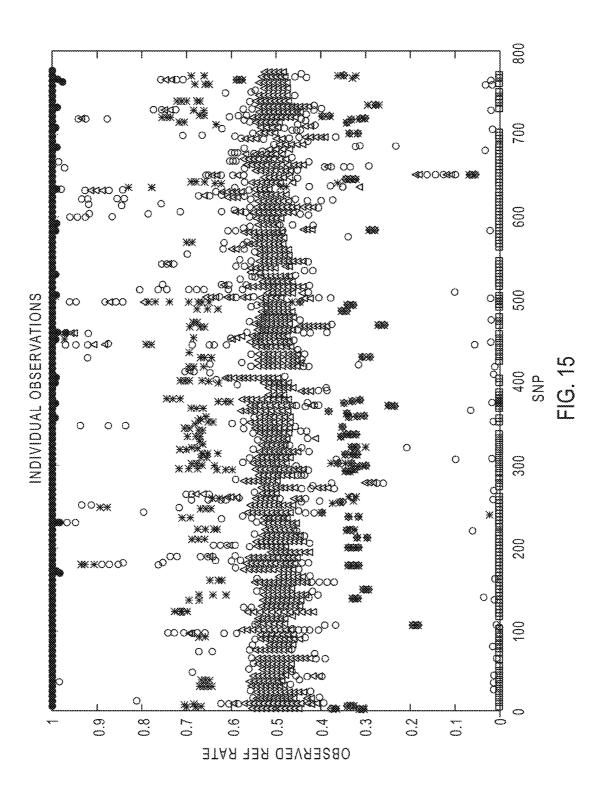
U.S. Pate	nt			Dec	c. 6,	202	2	Sheet 7 of 58			1	US	11,	519	,035 B2
SEQ ID NO:	44611	44612	44613	44614	44615	44616	******************************		44617	44618	44625	44620	44621	44622	
The sequencing-adaptor sequence is located inside the primer sequence and flanked by target specific sequence on both sides. 10 bases are target-specific at the 3'-end of each primer. Primers were tested successfully in real-time PCR. For sequencing this reduces the number of primer bases that need to be sequenced.	AACTCACATAGCACACGCTCTTCCGATCTTGCAAGCACA	TCCTCTGTGACACGCTCTTCCGATCTCCCTGCTCTT	toctotototACGACGCTCTTCCGATCTcGGGCTGTCA	TACATCCTTGAGACACGACGCTCTTCCGATCTGCTGTGCAGT	tttgcttgagctACACGACGCTCTTCCGATCTcgggagtttc	gtottatggtggACACGACGCTCTTCCGATCTcaaagccagt		The sequencing-adaptor sequence is located inside the primer sequence and flanked by target specific sequence on both sides. The internal tag is formed into a hairpin structure by 10 complementary bases on either end. This brings the target-specific ends of the primer into close proximity and hinders unspecific binding to the "internal tag". 10 bases are target-specific at the 3'-end of each primer. Primers were tested successfully in real-time PCR.	AACTCACATAGC tgatcggtACACGACGCTCTTCCGATCTTGCAAGCACA	TCCTCTGTG tgatcggtACACGACGCTCTTCCGATCTCCCTGCTCTT	tcotctcttgatcggtACACGACGCTCTTCCGATCTCGGGCTGTCA	TACATCCTTGAGtgatcggtACACGACGCTCTTCCGATCTGCTGTGCAGT	tttgottgagottgatcggtACACGACGCTCTTCCGATCTcgggagtttc	gtottatggtggtgatcggtACACGACGCTCTTCCGATCTcaaagccagt	FG. 12
ng-adaptor sequence ) bases are target-spi ng this reduces the m	int-tag 1.10	int-tag 2.10	int-tag 3.10	int-tag 4,10	Int-tag 5,10	int-tag 6,10		ng-adaptor sequence in internal tag is form sends of the primer is at the 3'-end of each	loop-int-tag 1.10	loop-int-tag 2.10	loop-int-tag 3.10	loop-int-tag 4.10	loop-int-tag 5.10	loop-int-tag 6.10	
The sequenci both sides. 10 For sequencir	rs8130564	rs2832093	rs12011281	rs6719561	rs10187018	rs10460481		The sequenci both sides. The target-specific target-specific	rs8130564	rs2832093	rs12011281	rs6719561	rs10187018	rs10460481	

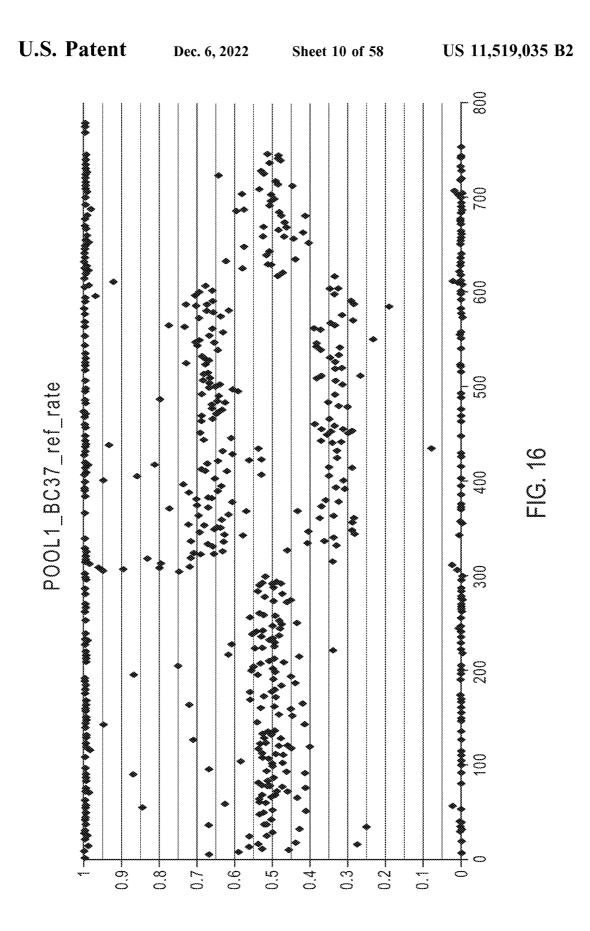
U.S. Patent Dec. 6, 2022 Sheet 8 of 58 US 11,519,035 B2



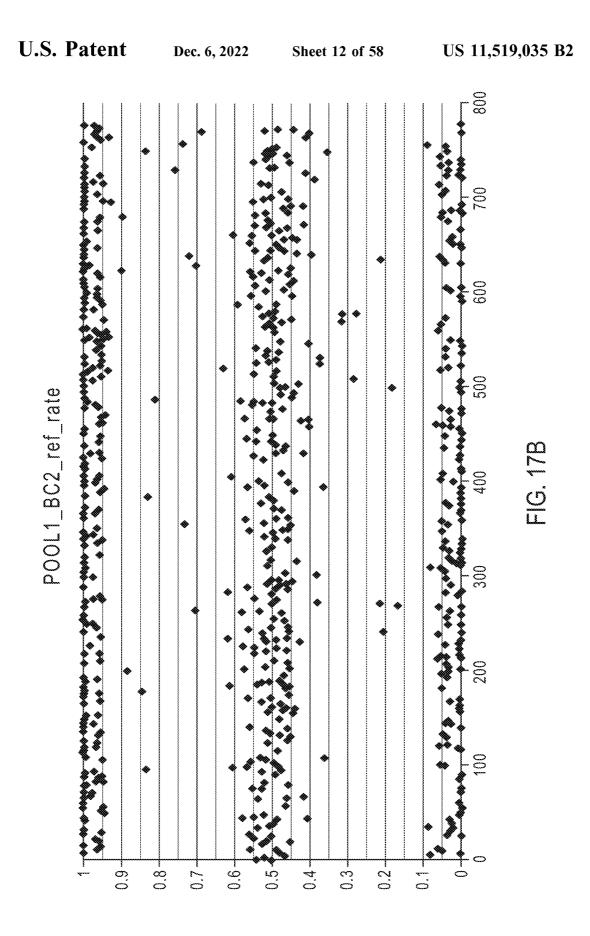


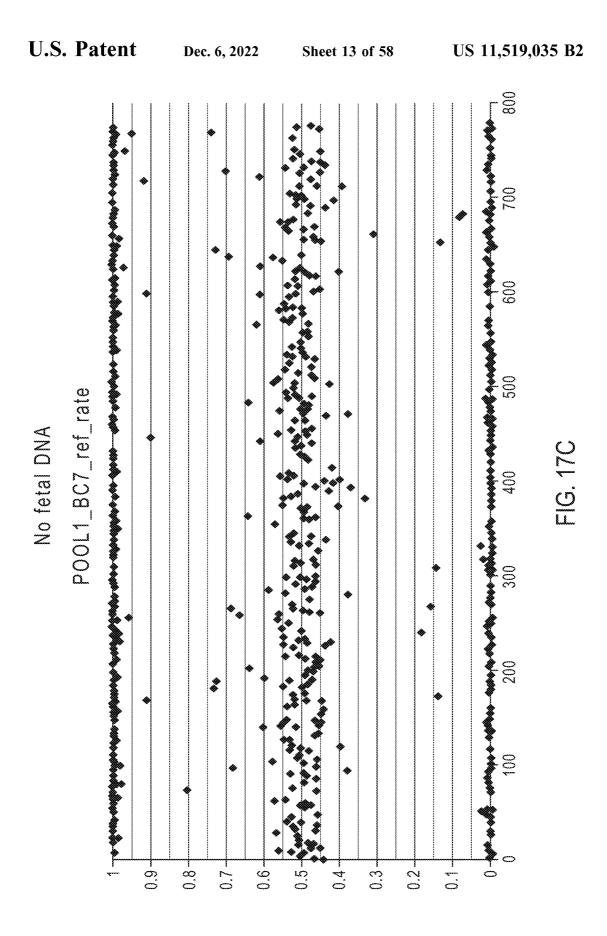
U.S. Patent Dec. 6, 2022 Sheet 9 of 58 US 11,519,035 B2

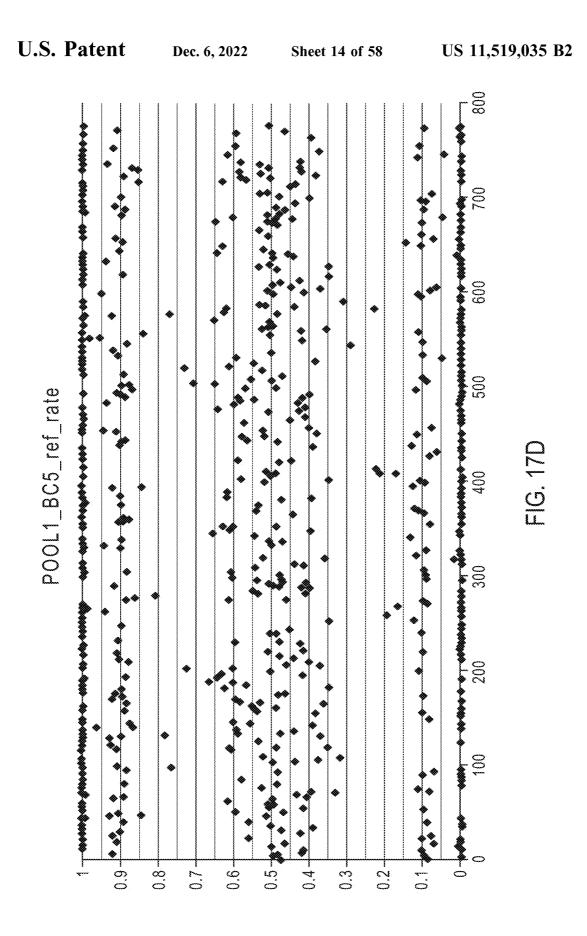




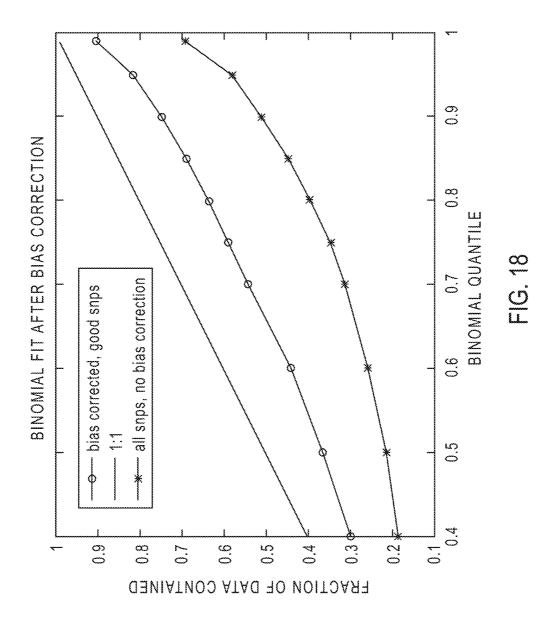
U.S. Patent Dec. 6, 2022 Sheet 11 of 58 US 11,519,035 B2 700 900 500 POOL1\_BC1\_ref\_rate 以 <u>で</u> <u>で</u> 200



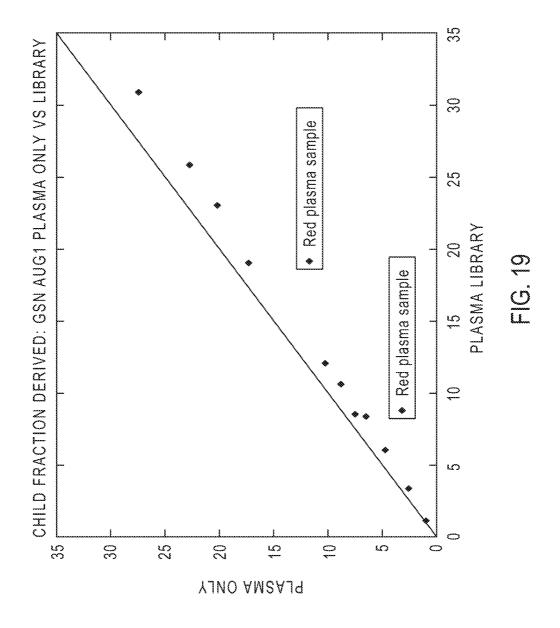




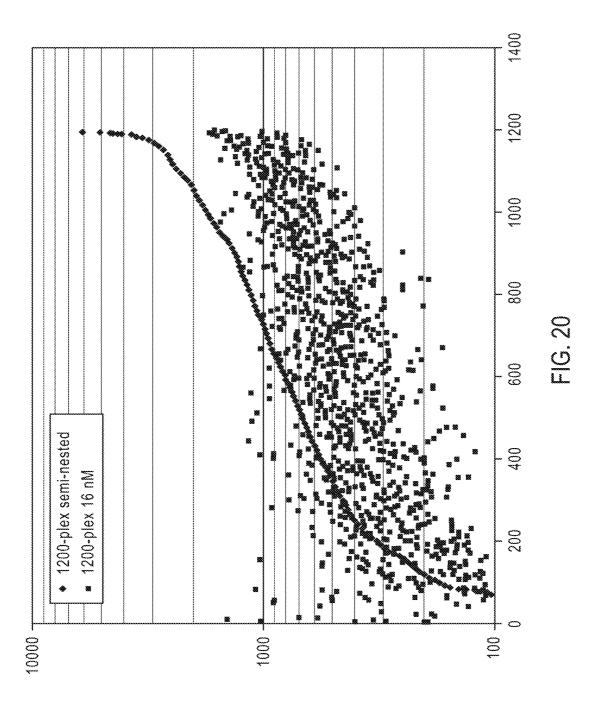
U.S. Patent Dec. 6, 2022 Sheet 15 of 58 US 11,519,035 B2

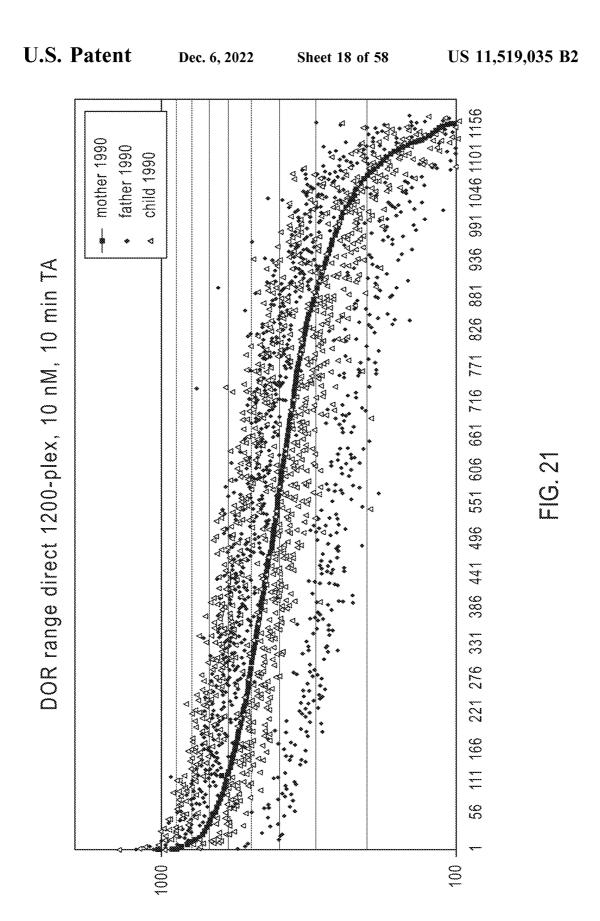


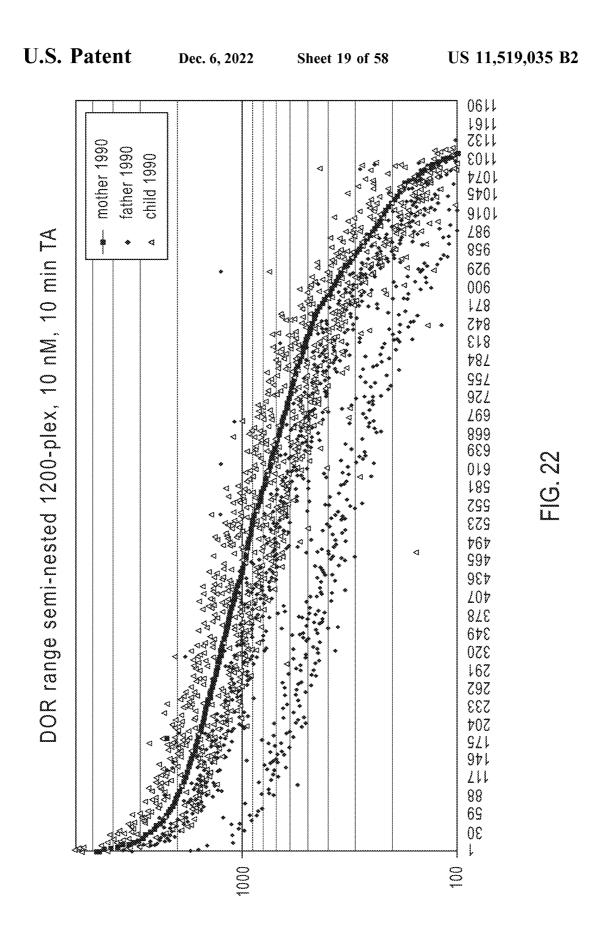
U.S. Patent Dec. 6, 2022 Sheet 16 of 58 US 11,519,035 B2



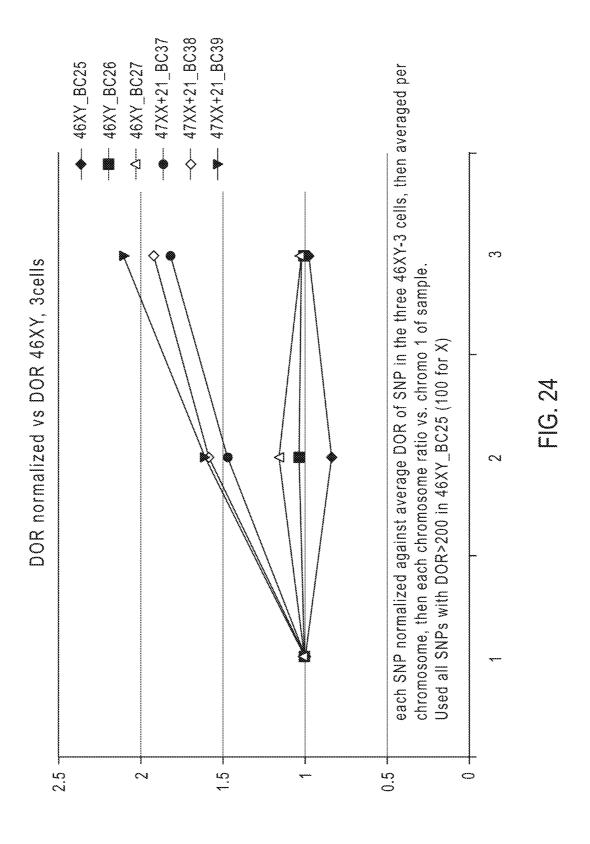
U.S. Patent Dec. 6, 2022 Sheet 17 of 58 US 11,519,035 B2

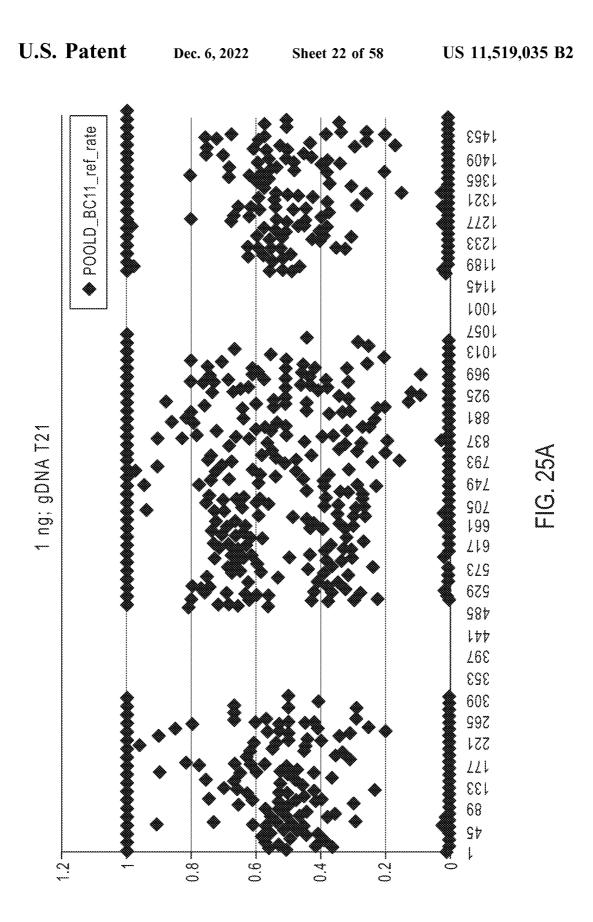


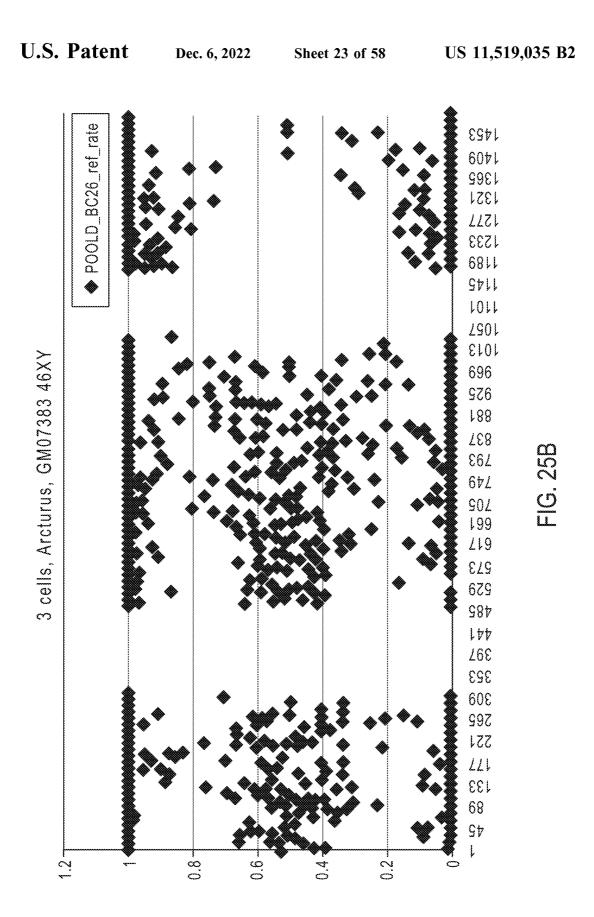




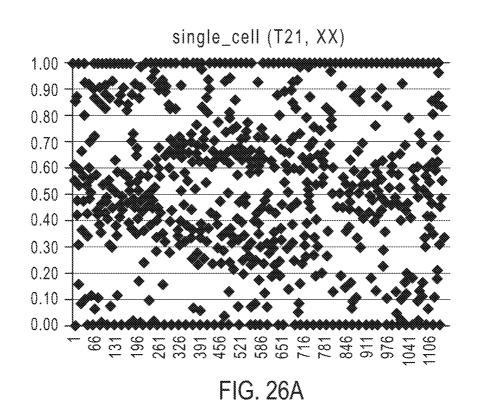
U.S. Patent Dec. 6, 2022 Sheet 21 of 58 US 11,519,035 B2

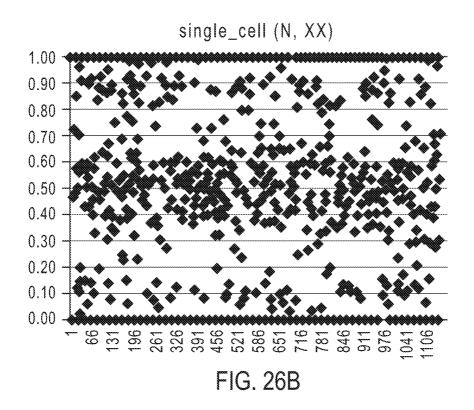




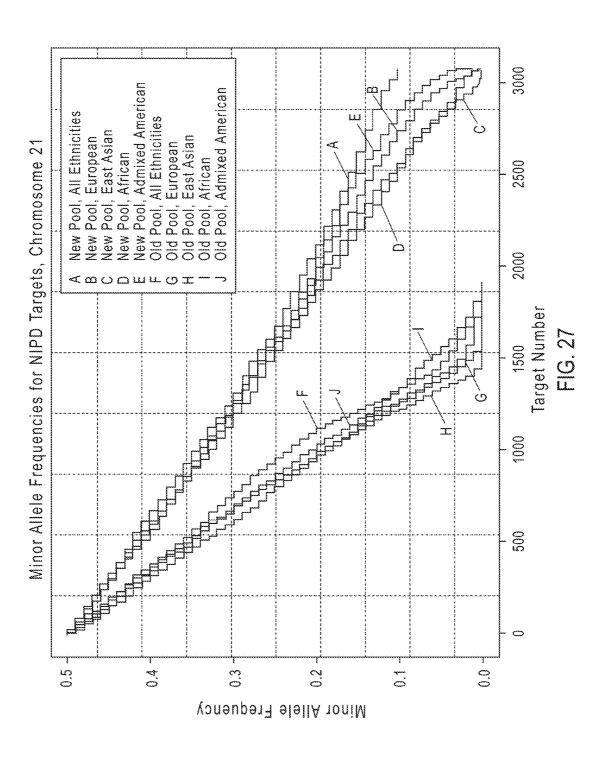


U.S. Patent Dec. 6, 2022 Sheet 25 of 58 US 11,519,035 B2

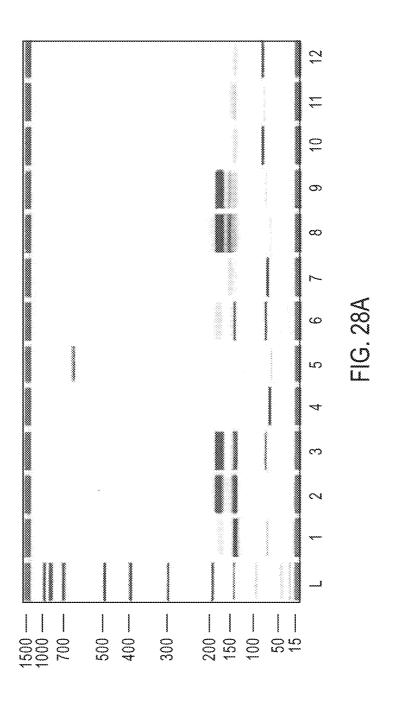


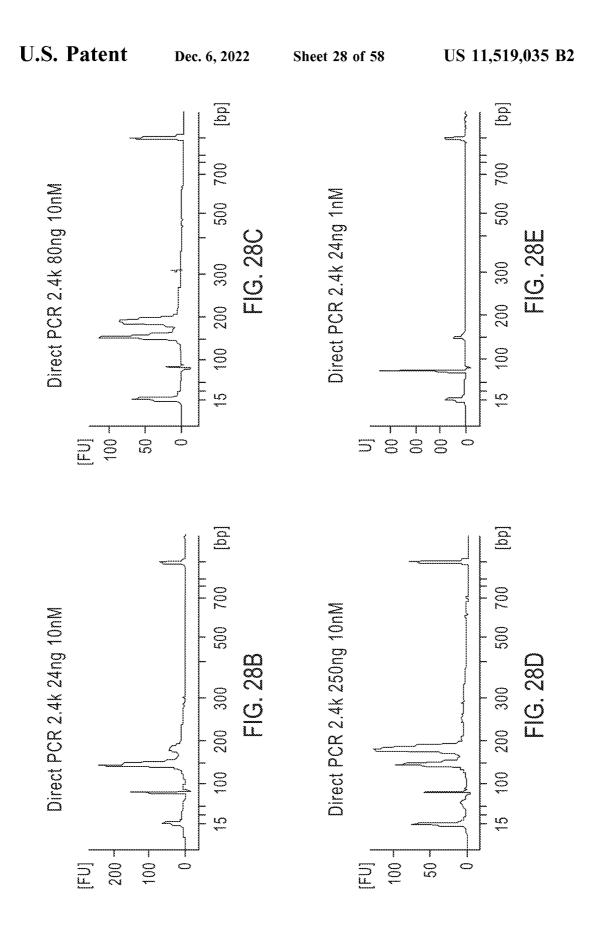


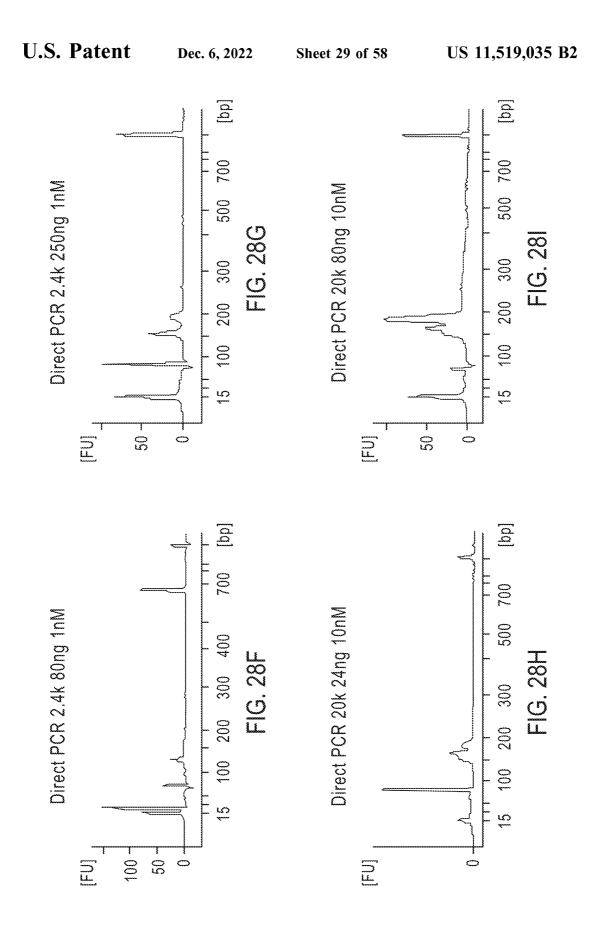
U.S. Patent Dec. 6, 2022 Sheet 26 of 58 US 11,519,035 B2

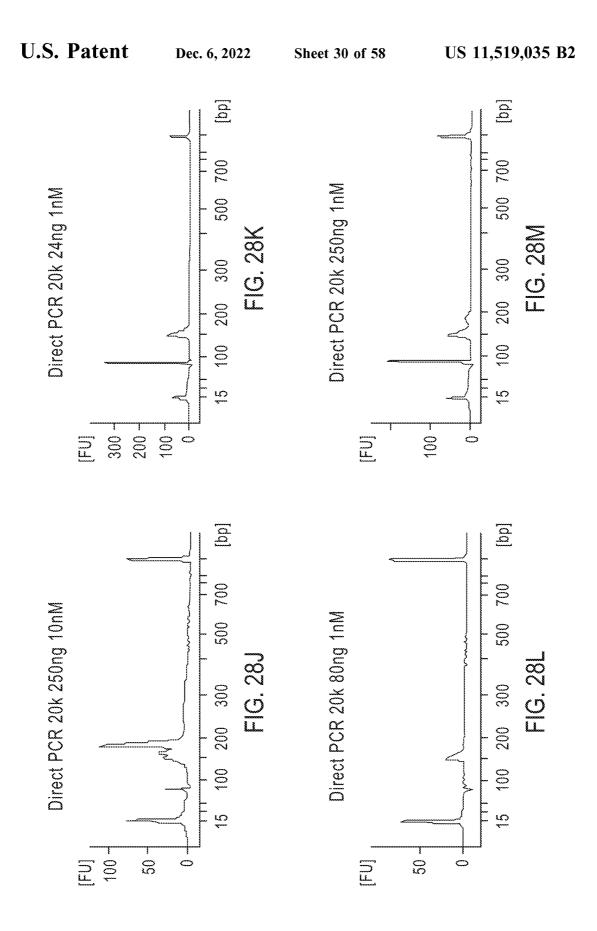


U.S. Patent Dec. 6, 2022 Sheet 27 of 58 US 11,519,035 B2

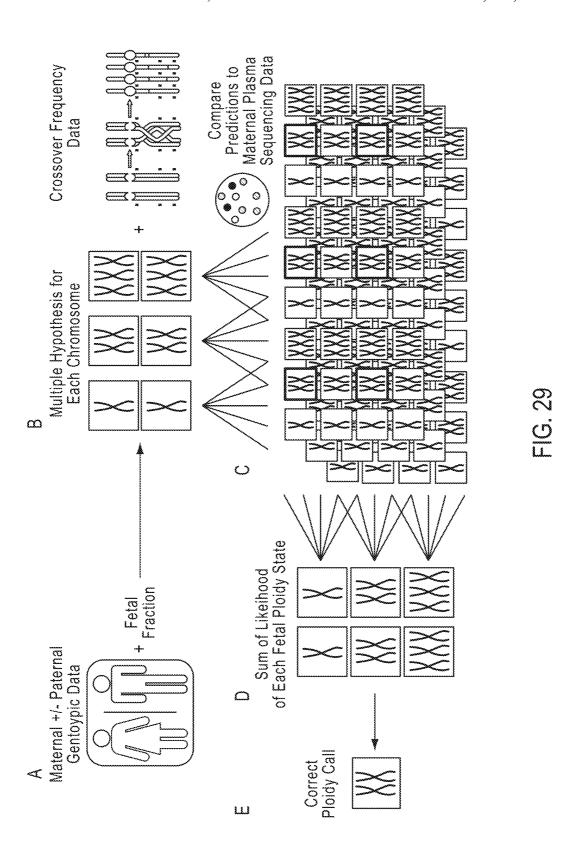


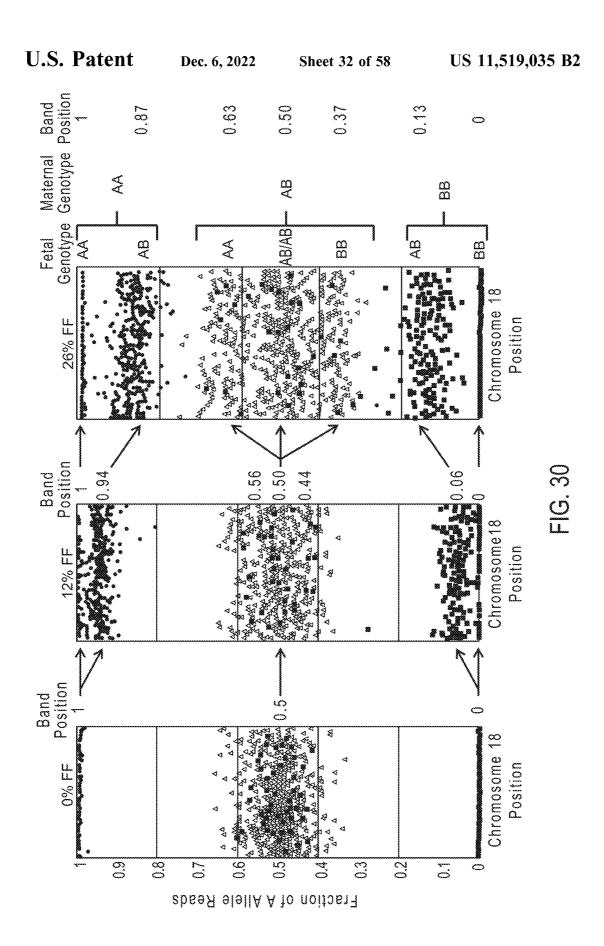


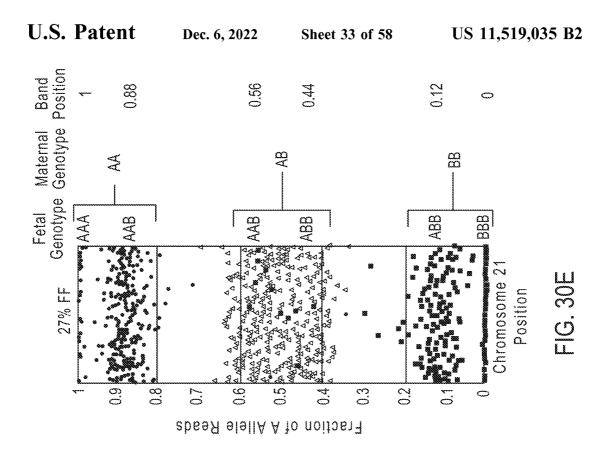


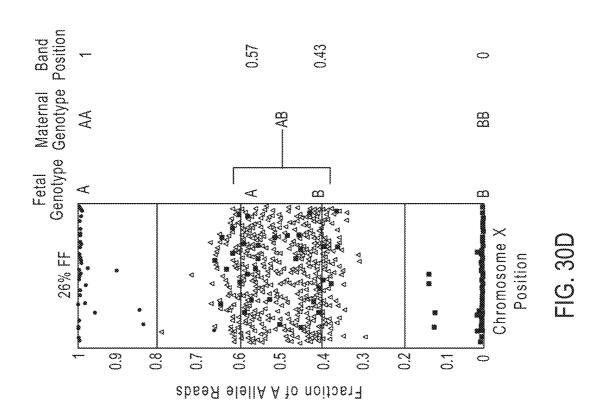


U.S. Patent Dec. 6, 2022 Sheet 31 of 58 US 11,519,035 B2

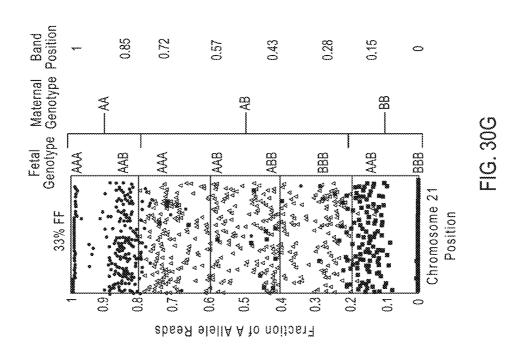


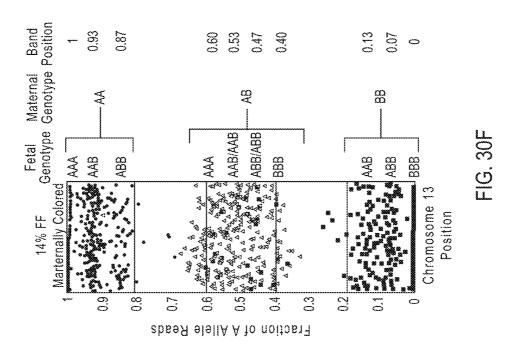






U.S. Patent Dec. 6, 2022 Sheet 34 of 58 US 11,519,035 B2





U.S. Patent Dec. 6, 2022 Sheet 35 of 58 US 11,519,035 B2

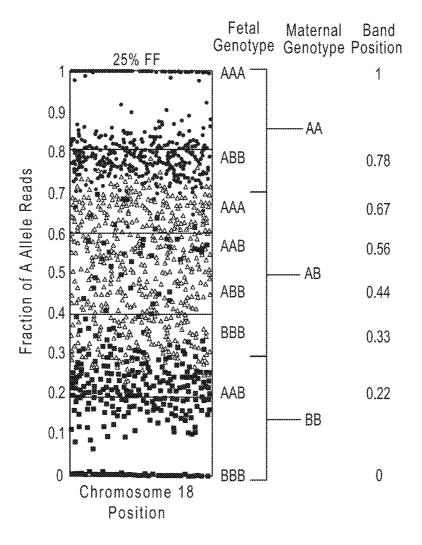
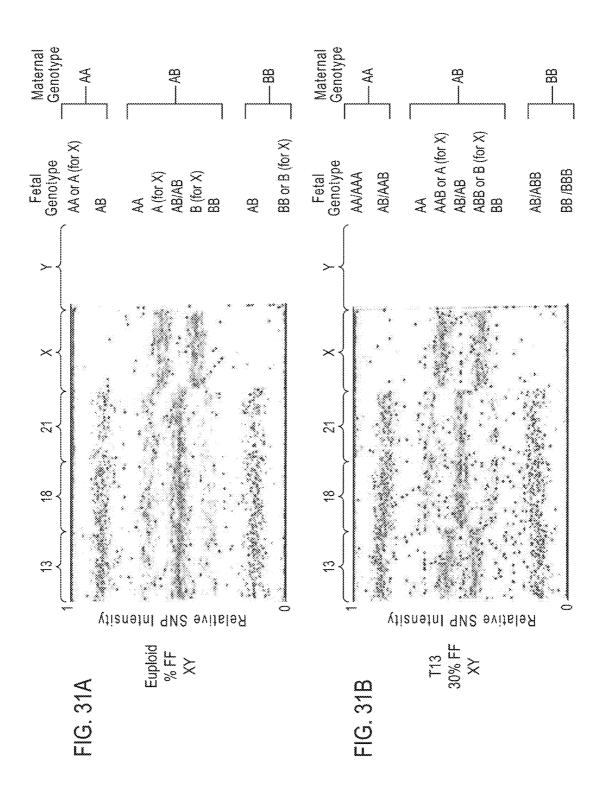
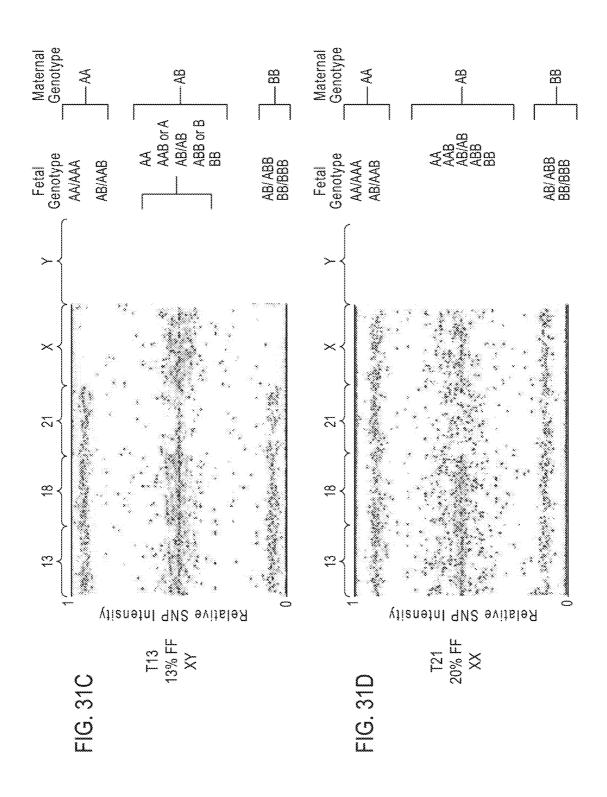


FIG. 30H

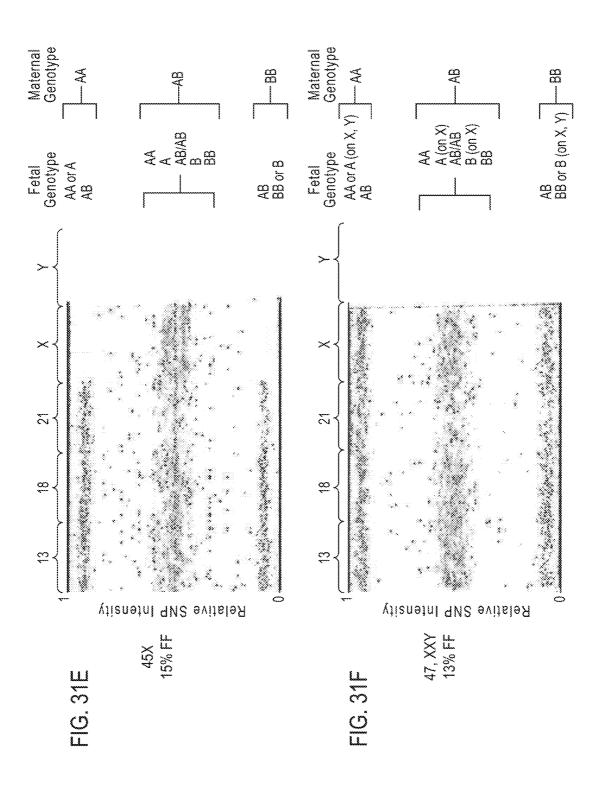
U.S. Patent Dec. 6, 2022 Sheet 36 of 58 US 11,519,035 B2



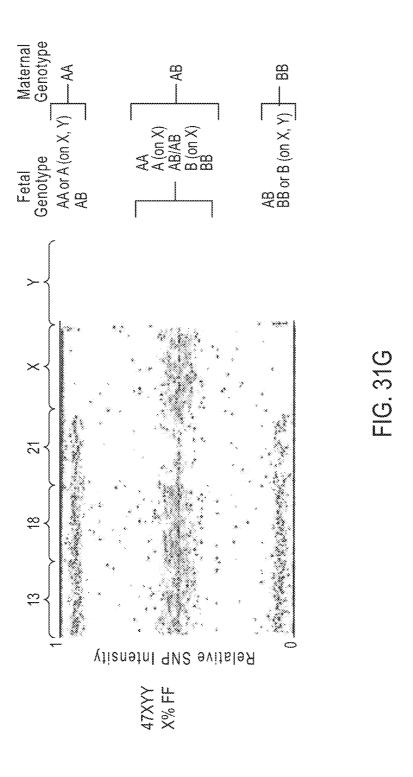
U.S. Patent Dec. 6, 2022 Sheet 37 of 58 US 11,519,035 B2



U.S. Patent Dec. 6, 2022 Sheet 38 of 58 US 11,519,035 B2



U.S. Patent Dec. 6, 2022 Sheet 39 of 58 US 11,519,035 B2



U.S. Patent Dec. 6, 2022 Sheet 40 of 58 US 11,519,035 B2

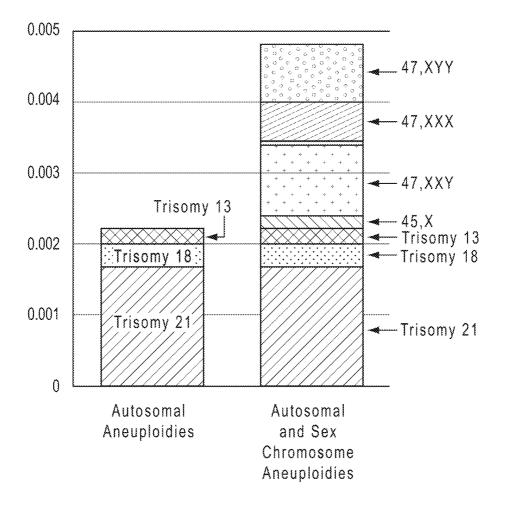


FIG. 32

U.S. Patent Dec. 6, 2022 Sheet 41 of 58 US 11,519,035 B2

FIG. 33A

FIG. 33B

U.S. Patent Dec. 6, 2022 Sheet 42 of 58 US 11,519,035 B2

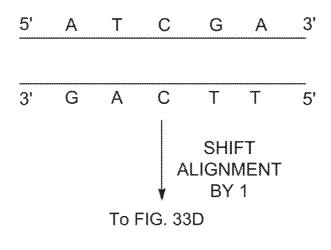


FIG. 33C



U.S. Patent Dec. 6, 2022 Sheet 43 of 58 US 11,519,035 B2

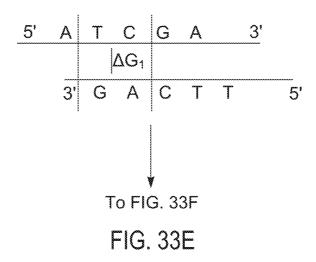


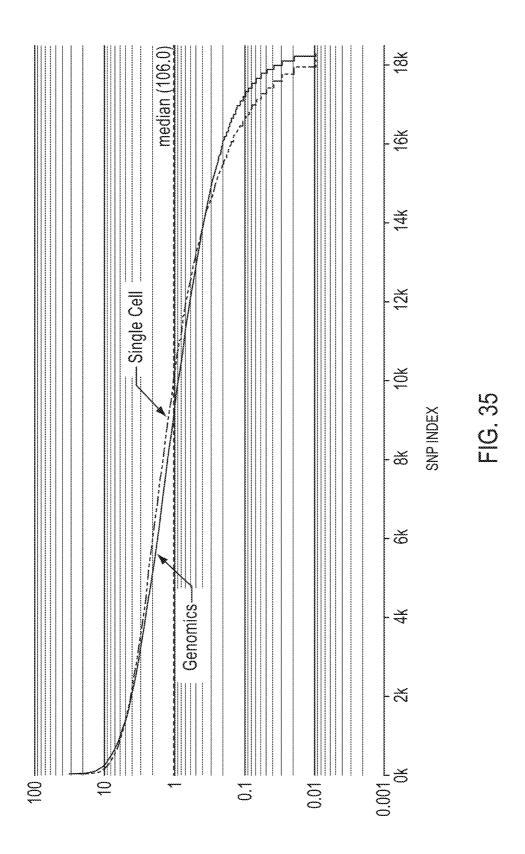
FIG. 33F

U.S. Patent Dec. 6, 2022 Sheet 44 of 58 US 11,519,035 B2

Description	MAPPED_READS   TOTAL_READS   %Mapped Reads	TOTAL_READS	%Wapped Read
gMother	3,633,900	3,647,446	100%
gMother	3,567,685	3,578,549	100%
gMother	3,733,706	3,747,692	100%
gChild	3,725,554	3,737,742	700%
gChild	3,816,989	3,829,450	700%
gChild	3,524,956	3,536,569	700%
1 mother	3,467,493	7,564,869	46%
1 mother	3,868,210	7,107,028	54%
1 mother	4,918,140	7,613,240	%59
1 child	361,390	6,507,434	%9
1 child	1,885,864	6,496,348	79%
1 child	2,789,647	6,259,288	45%

五 (2) (2)

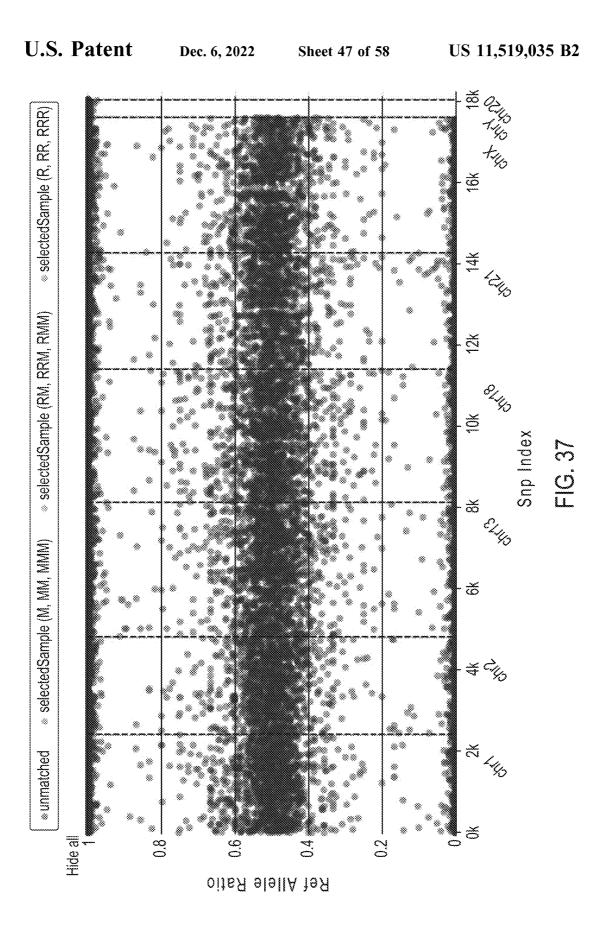
U.S. Patent Dec. 6, 2022 Sheet 45 of 58 US 11,519,035 B2



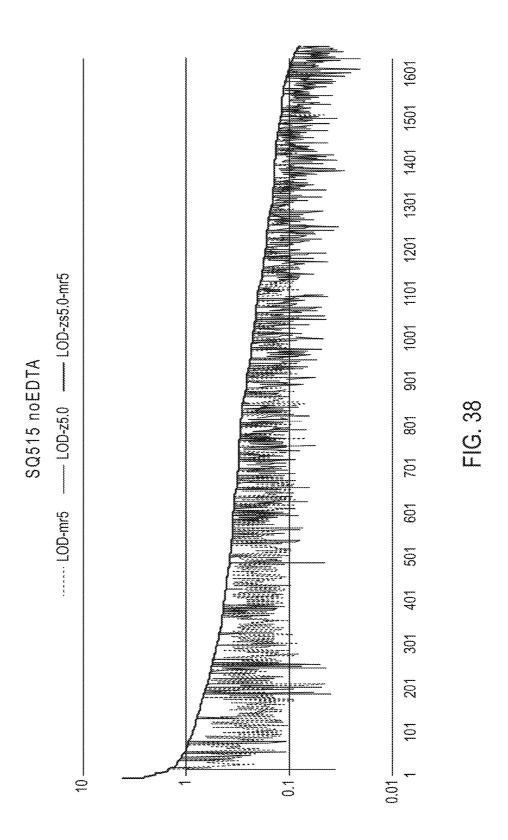
U.S. Patent Dec. 6, 2022 Sheet 46 of 58 US 11,519,035 B2

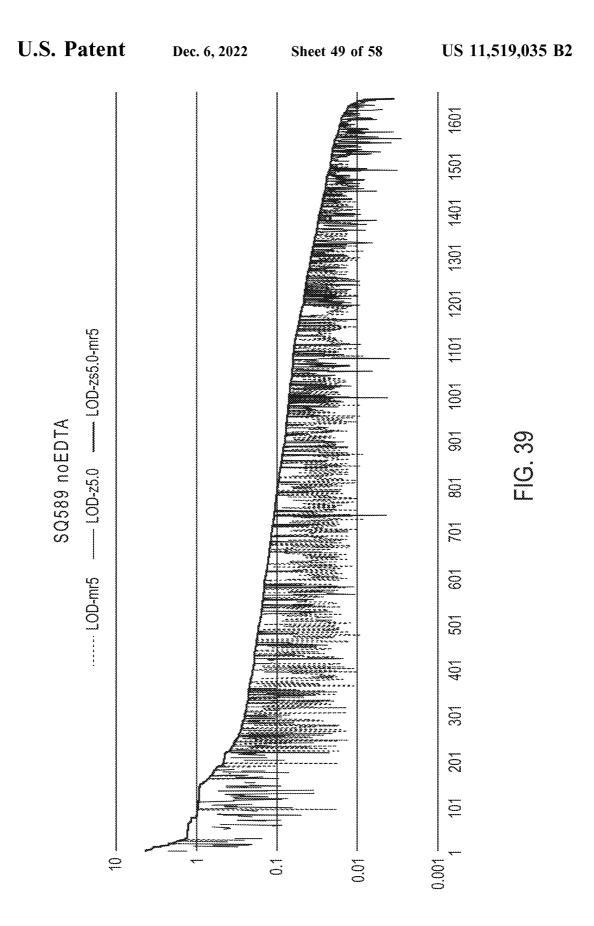
Description	MAPPED_READS	TOTAL_READS	MAPPED_READS   TOTAL_READS   %Mapped Reads
case 1 blastoceol	68,123	8,470,872	1%
case 1 cell 1	4,340,613	8,248,598	53%
case 1 cell 2	5,480,580	8,230,870	67%
case 1 cell 3	4,664,577	7,846,040	29%
case 2 blastoceol	45,794	6,302,957	1%
case 2 cell 1	6,988,854	8,622,495	81%
case 2 cell 2	7,083,600	8,843,495	80%
case 2 cell 3	5,811,364	8,256,310	70%

FG. 38

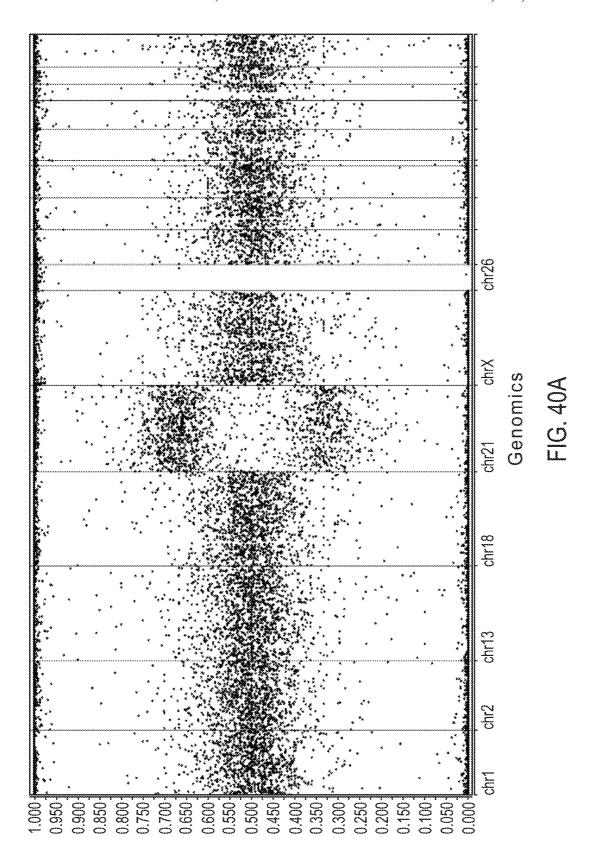


U.S. Patent Dec. 6, 2022 Sheet 48 of 58 US 11,519,035 B2

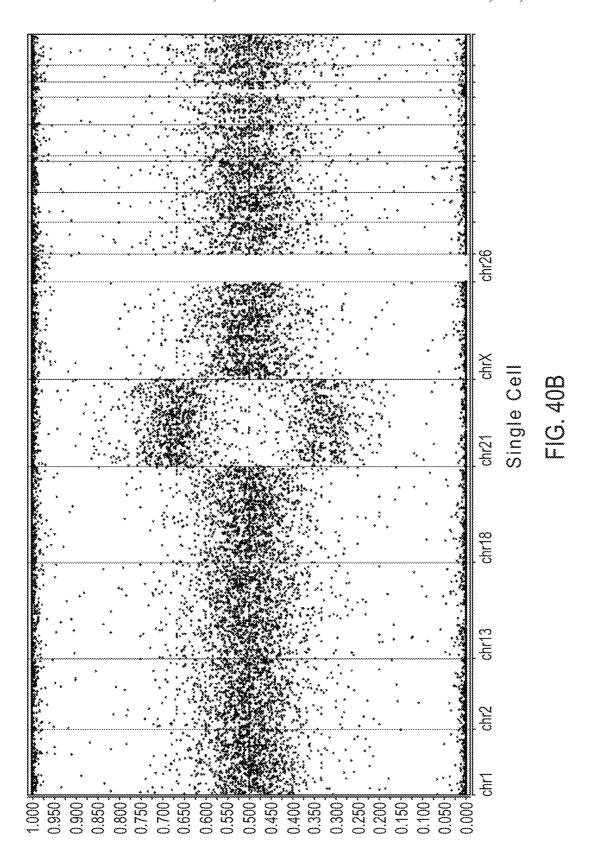




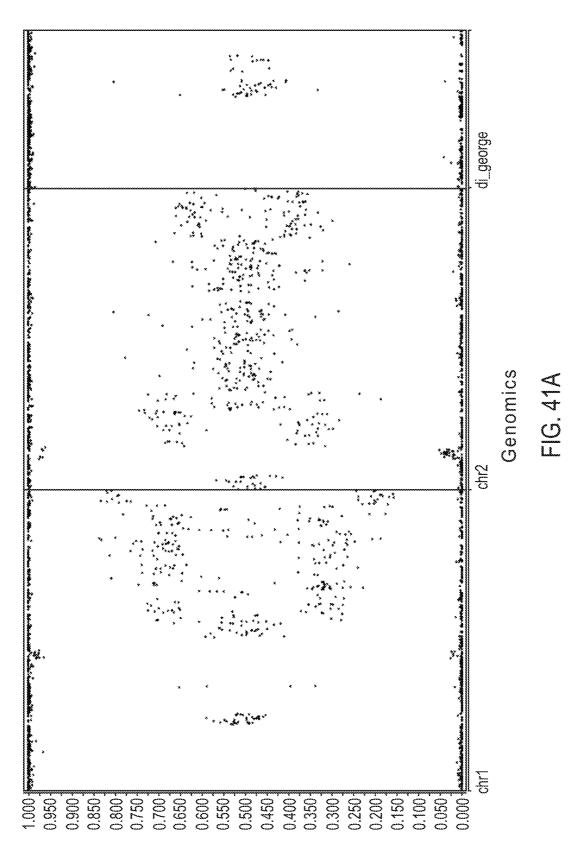
U.S. Patent Dec. 6, 2022 Sheet 50 of 58 US 11,519,035 B2



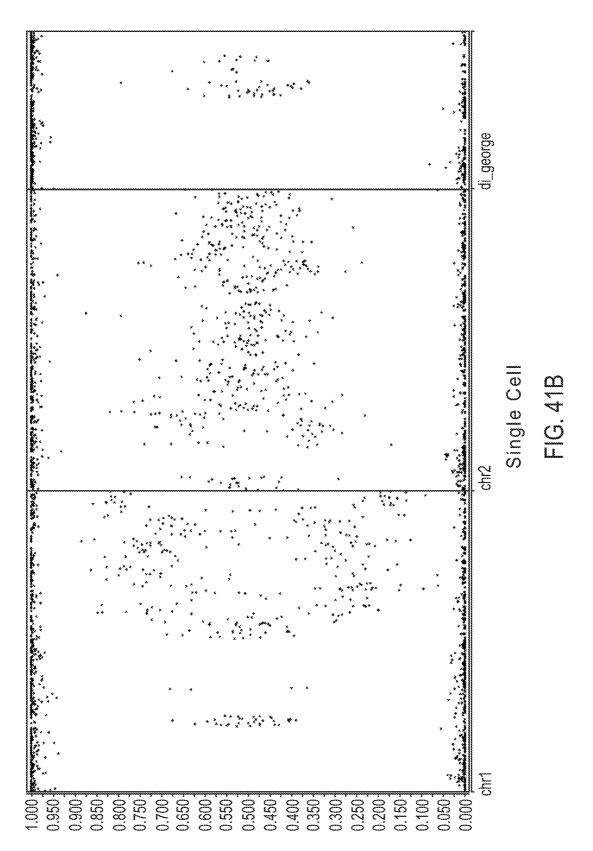
U.S. Patent Dec. 6, 2022 Sheet 51 of 58 US 11,519,035 B2

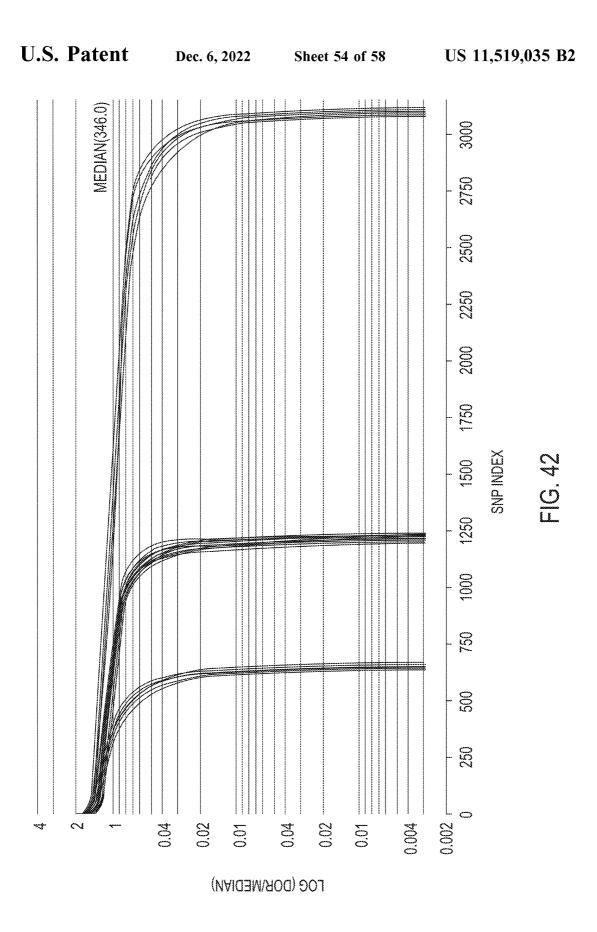


U.S. Patent Dec. 6, 2022 Sheet 52 of 58 US 11,519,035 B2



U.S. Patent Dec. 6, 2022 Sheet 53 of 58 US 11,519,035 B2

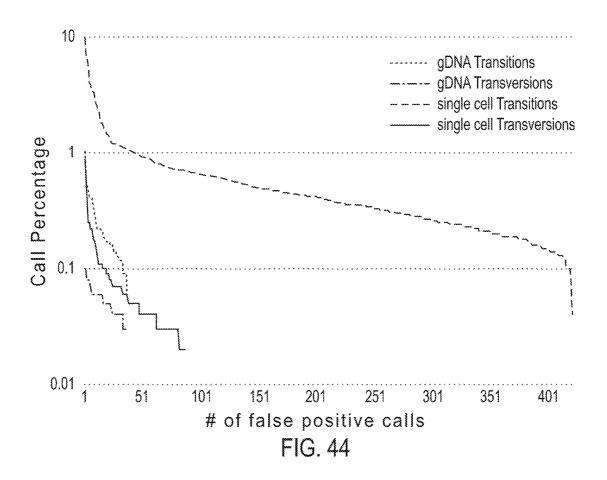




U.S. Patent Dec. 6, 2022 Sheet 55 of 58 US 11,519,035 B2

	gDNA	Single Cell
Count	75	510
Mean	0.15 %	0.51 %
Median	0.09 %	0.33 %
Max	1.03 %	10 %
Standard Deviation	0.16%	0.79 %
95 <sup>th</sup> percentile	0.43 %	1.22 %
90 <sup>th</sup> percentile	0.37 %	0.92 %

FIG. 43

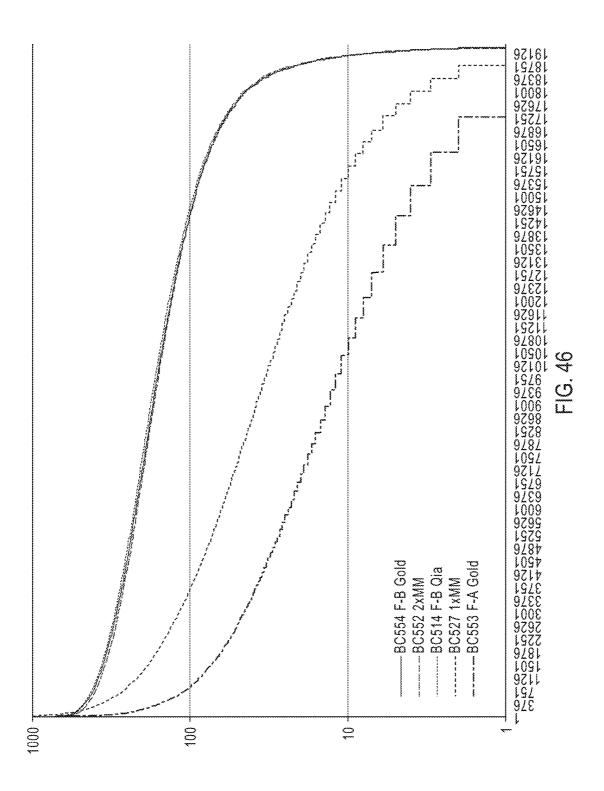


U.S. Patent Dec. 6, 2022 Sheet 56 of 58 US 11,519,035 B2

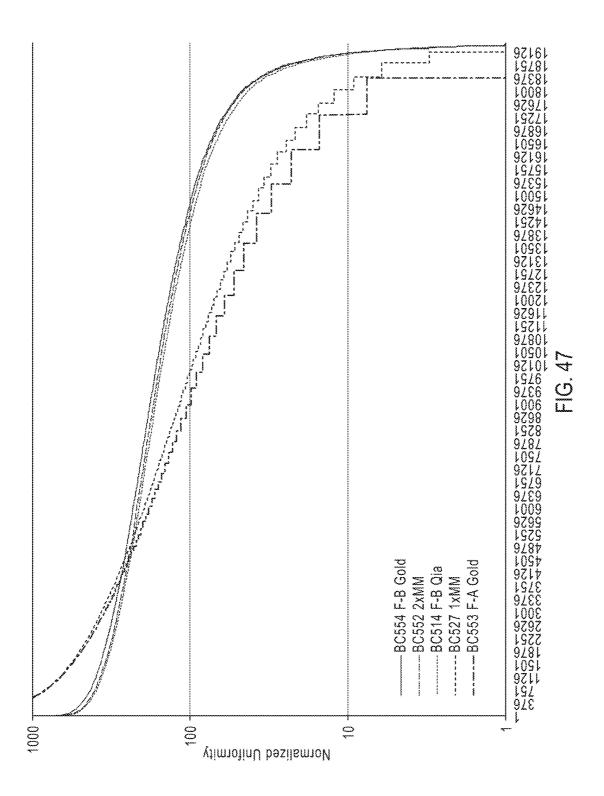
							 		**********		
	%,	%	%	%:	%	%1		aka F-A Gold	aka F-B Gold		naka F-B Qiagen
Error Rate (%)	0.107%	0.141%	0.077%	0.072%	0.064%	0.079%		ìold	Sold	Gold	en HS Tac
NOR	1,043,539	3,206,478	421,228	3,106,520	3,796,213	3,461,304		50 U/mL Taq G	F-B is : 75 mM Tris pH 7.8, 6 mM MgCl $_2$ , 0 mM KCl, 40 mM (NH $_4$ ) $_2$ SO $_4$ , 150 U/mL Taq Gold	8.2, 3 mM MgCl <sub>2</sub> , 30 mM KCl, 40 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 150 U/mL Taq Gold	7.8, 6 mM MgCl $_2$ , 0 mM KCl, 40 mM (NH $_4$ ) $_2$ SO $_4$ , 150 U/mL <b>Qiagen HS Taq</b> aka F-B Qiagen
Median DOR All	26.25	147.75	9.625	142.5	142.5	159.25		H <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> ,			
Dropout Count	436	105	1379	105	122	102		I, 10 mM (N			
% Mapped	80.0%	94.4%	70.6%	93.7%	98.0%	93.7%		2, 0 mM KC			
TOTAL READS	1,262,558	3,615,347	635,571	3,519,378	4,109,448	3,922,330		$7.8,3\mathrm{mM}\mathrm{MgCl}_2,$ 0 mM KCl, 10 mM (NH $_4)_2$ SO $_4$ , 50 U/mL Taq Gold			
Mapped READS	1,112,007	3,412,593	449,074	3,293,944	4,028,128	3,676,617		F-A is : 25 mM Tris pH 7		F-D is : 25 mM Tris <b>pH (</b>	F-J is : 75 mM Tris pH 7
D0E1	1xMM	2xMM	F-A	유 <u>-</u>	F-D	7.4		F-A is:2	F-B is : 7	F-D is: 2	F-J is : 7

而 (2)

U.S. Patent Dec. 6, 2022 Sheet 57 of 58 US 11,519,035 B2



U.S. Patent Dec. 6, 2022 Sheet 58 of 58 US 11,519,035 B2



Case: 24-1324 Document: 42-1 Page: 123 Filed: 03/18/2024

## US 11,519,035 B2

#### METHODS FOR SIMULTANEOUS AMPLIFICATION OF TARGET LOCI

#### CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. Utility application Ser. No. 16/743,724, filed Jan. 15, 2020.

U.S. Utility application Ser. No. 16/743,724 is a continuation of U.S. Utility application Ser. No. 16/399,268 (now 10 U.S. Pat. No. 10,538,814), filed Apr. 30, 2019.

U.S. Utility application Ser. No. 16/399,268 (now U.S. Pat. No. 10,538,814) is a continuation of U.S. Utility application Ser. No. 16/140,298, filed Sep. 24, 2018.

U.S. Utility application Ser. No. 16/140,298 is a continu- 15 ation of U.S. Utility application Ser. No. 14/918,544 (now U.S. Pat. No. 10,316,362), filed Oct. 20, 2015.

U.S. Utility application Ser. No. 14/918,544 (now U.S. Pat. No. 10,316,362), filed Oct. 20, 2015, is a continuationin-part application of U.S. Utility application Ser. No. 20 14/877,925, filed Oct. 7, 2015, now abandoned; a continuation-in-part application of U.S. Utility application Ser. No. 14/692,703, filed Apr. 21, 2015, now U.S. Pat. No. 10,179, 937; a continuation-in-part application of U.S. Utility application Ser. No. 14/538,982, now U.S. Pat. No. 9,677,118, 25 filed Nov. 24, 2014; and claims the benefit of and priority to U.S. Provisional Application Ser. No. 62/148,173, filed Apr. 15, 2015; U.S. Provisional Application Ser. No. 62/147,377, filed Apr. 14, 2015; U.S. Provisional Application Ser. No. 62/146,188, filed Apr. 10, 2015; and U.S. Provisional Appli- 30 cation Ser. No. 62/066,514, filed Oct. 21, 2014.

U.S. Utility application Ser. No. 14/877,925, filed Oct. 7, 2015, now abandoned, is a continuation-in-part of U.S. Utility application Ser. No. 14/225,356 (now Abandoned), filed Mar. 25, 2014; is a continuation-in-part of U.S. Utility application Ser. No. 13/780,022 (now Abandoned), filed Feb. 28, 2013; and is a continuation of U.S. Utility application Ser. No. 13/683,604 (now Abandoned), filed Nov. 21,

 $U.S.\ Utility\ application\ Ser.\ No.\ 14/692,703,\ filed\ Apr.\ 21,\ \ 40$ 2015, now U.S. Pat. No. 10,179,937, claims the benefit of and priority to U.S. Provisional Application Ser. No. 62/148, 173, filed Apr. 15, 2015; U.S. Provisional Application Ser. No. 62/147,377, filed Apr. 14, 2015; U.S. Provisional Application Ser. No. 62/146,188, filed Apr. 10, 2015; U.S. Pro- 45 visional Application Ser. No. 62/066,514, filed Oct. 21, 2014; U.S. Provisional Application Ser. No. 61/994,791, filed May 16, 2014; U.S. Provisional Application Ser. No. 61/987,407, filed May 1, 2014; and U.S. Provisional Application Ser. No. 61/982,245, filed Apr. 21, 2014.

U.S. Utility application Ser. No. 14/538,982, now U.S. Pat. No. 9,677,118, filed Nov. 24, 2014 claims the benefit of and priority to U.S. Provisional Application Ser. No. 62/066, 514, filed Oct. 21, 2014; U.S. Provisional Application Ser. No. 61/994,791, filed May 16, 2014; U.S. Provisional Appli- 55 and is 8,726 kilo bytes in size. cation Ser. No. 61/987,407, filed May 1, 2014; and U.S. Provisional Application Ser. No. 61/982,245, filed Apr. 21,

U.S. Utility application Ser. No. 14/225,356 (now Aban-Application PCT/US2012/58578, filed Oct. 3, 2012.

U.S. Utility application Ser. No. 13/780,022 (now Abandoned), filed Feb. 28, 2013, is a continuation-in-part of U.S. Utility application Ser. No. 13/683,604 (now Abandoned), filed Nov. 21, 2012; a continuation-in-part of PCT Applica- 65 tion No. PCT/US2012/58578, filed Oct. 3, 2012; a continuation-in-part of U.S. Utility application Ser. No. 13/335,043,

filed Dec. 22, 2011, now U.S. Pat. No. 10,113,196; a continuation-in-part of U.S. Utility application Ser. No. 13/300,235, filed Nov. 18, 2011, now U.S. Pat. No. 10,017, 812; and an continuation-in-part of U.S. Utility application Ser. No. 13/110,685 (now U.S. Pat. No. 8,825,412), filed May 18, 2011, and also claims the benefit of U.S. Provisional Application Ser. No. 61/634,431, filed Feb. 29, 2012.

U.S. Utility application Ser. No. 13/683,604 (now Abandoned), filed Nov. 21, 2012, is a continuation-in-part of U.S. Utility application Ser. No. 13/300,235 (now U.S. Pat. No. 10,017,812), filed Nov. 18, 2011; is a continuation-in-part of U.S. Utility application Ser. No. 13/110,685 (now U.S. Pat. No. 8,825,412), filed May 18, 2011; and claims the benefit of and priority to U.S. Provisional Application Ser. No. 61/675,020, filed Jul. 24, 2012.

PCT Application No. PCT/US2012/058578, filed Oct. 3, 2012, is a continuation-in-part of U.S. Utility application Ser. No. 13/300,235 (now U.S. Pat. No. 10,017,812), filed Nov. 18, 2011; and claims the benefit of and priority to U.S. Provisional Application Ser. No. 61/683,331, filed Aug. 15, 2012; and U.S. Provisional Application Ser. No. 61/542,508, filed Oct. 3, 2011.

U.S. Utility application Ser. No. 13/335,043, filed Dec. 22, 2011, is a continuation-in-part of U.S. Utility application Ser. No. 13/300,325 (now U.S. Pat. No. 10,017,812), filed Nov. 18, 2011; a continuation-in-part of U.S. Utility application Ser. No. 13/110,685 (now U.S. Pat. No. 8,825,412), filed May 18, 2011; and claims the benefit of U.S. Provisional Application Ser. No. 61/426,208, filed Dec. 22, 2010.

U.S. Utility application Ser. No. 13/300,235 (now U.S. Pat. No. 10,017,812), filed Nov. 18, 2011 is a continuationin-part of U.S. Utility application Ser. No. 13/110,685 (now U.S. Pat. No. 8,825,412), filed May 18, 2011 and claims the benefit of U.S. Provisional Application Ser. No. 61/542,508, filed Oct. 3, 2011; and U.S. Provisional Application Ser. No. 61/571,248, filed Jun. 23, 2011.

U.S. Utility application Ser. No. 13/110,685 (now U.S. Pat. No. 8,825,412), filed May 18, 2011, claims the benefit of U.S. Provisional Application Ser. No. 61/516,996, filed Apr. 12, 2011; U.S. Provisional Application Ser. No. 61/448, 547, filed Mar. 2, 2011; U.S. Provisional Application Ser. No. 61/462,972, filed Feb. 9, 2011; U.S. Provisional Application Ser. No. 61/398,159, filed Jun. 21, 2010; and U.S. Provisional Application Ser. No. 61/395,850, filed May 18,

Each of these applications cited above is hereby incorporated by reference in its entirety.

## SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Jul. 8, 2020, is named N012US32\_SL.txt

## FIELD OF THE INVENTION

The present invention generally relates to methods and doned), filed Mar. 25, 2014 is a continuation of PCT 60 compositions for simultaneously amplifying multiple nucleic acid regions of interest in one reaction volume.

#### BACKGROUND OF THE INVENTION

To increase assay throughput and allow more efficient use of nucleic acid samples, simultaneous amplification of many target nucleic acids in a sample of interest can be carried out

## US 11,519,035 B2

.

by combining many oligonucleotide primers with the sample and then subjecting the sample to polymerase chain reaction (PCR) conditions in a process known in the art as multiplex PCR. Use of multiplex PCR can significantly simplify experimental procedures and shorten the time required for nucleic acid analysis and detection. However, when multiple pairs are added to the same PCR reaction, non-target amplification products may be generating such products increases as the number of primers increases. These non-target amplicons significantly limit the use of the amplified products for further analysis and/or assays. Thus, improved methods are needed to reduce the formation of non-target amplicons during multiplex PCR.

Improved multiplex PCR methods would be useful for a 15 variety of application, such as Non-Invasive Prenatal Genetic Diagnosis (NPD). In particular, current methods of prenatal diagnosis can alert physicians and parents to abnormalities in growing fetuses. Without prenatal diagnosis, one in 50 babies is born with serious physical or mental handi- 20 cap, and as many as one in 30 will have some form of congenital malformation. Unfortunately, standard methods have either poor accuracy, or involve an invasive procedure that carries a risk of miscarriage. Methods based on maternal blood hormone levels or ultrasound measurements are non- 25 invasive, however, they also have low accuracies. Methods such as amniocentesis, chorion villus biopsy and fetal blood sampling have high accuracy, but are invasive and carry significant risks. Amniocentesis was performed in approximately 3% of all pregnancies in the US, though its frequency of use has been decreasing over the past decade and a half.

Normal humans have two sets of 23 chromosomes in every healthy, diploid cell, with one copy coming from each parent. Aneuploidy, a condition in a nuclear cell where the cell contains too many and/or too few chromosomes is believed to be responsible for a large percentage of failed implantations, miscarriages, and genetic diseases. Detection of chromosomal abnormalities can identify individuals or embryos with conditions such as Down syndrome, Klinefelter's syndrome, and Turner syndrome, among others, in addition to increasing the chances of a successful pregnancy. Testing for chromosomal abnormalities is especially important as the mother's age: between the ages of 35 and 40 it is estimated that at least 40% of the embryos are abnormal, and above the age of 40, more than half of the embryos are abnormal.

It has recently been discovered that cell-free fetal DNA and intact fetal cells can enter maternal blood circulation. Consequently, analysis of this genetic material can allow early NPD. Improved methods are desired to improve the 50 sensitivity and specificity and reduce the time and cost required for NPD.

#### SUMMARY OF THE INVENTION

In one aspect, the invention features methods of amplifying target loci in a nucleic acid sample. In some embodiments, the method involves (i) contacting the nucleic acid sample with a library of test primers (such as non-immobilized primers) that simultaneously hybridize to at least 25; 60; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci to produce a reaction mixture; and (ii) subjecting the reaction mixture to primer extension reaction conditions to 65 produce amplified products that include target amplicons. In some embodiments, the method also includes determining

4

the presence or absence of at least one target amplicon (such as at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target amplicons). In some embodiments, the method also includes determining the sequence of at least one target amplicon (such as at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target amplicons). In some embodiments, the method involves multiplex PCR and sequencing (such as high throughput sequencing). In some embodiments, the method includes selecting the test primers from a library of candidate primers by removing one or more of the candidate primers based at least in part on the likelihood of dimer formation between candidate primers (such as  $\Delta G$  values, undesirability scores, or interaction scores) prior to contacting the nucleic acid sample with the library of test primers.

In some embodiments, the method involves (i) contacting a sample comprising target human loci with a library of at least 50 (such as at least 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000) non-immobilized, non-identical primers that simultaneously hybridize to at least 50 (such as at least 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000) non-identical target human loci to produce a reaction mixture; wherein the primers do not include molecular inversion probes (MIPs); (ii) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products comprising target amplicons; wherein the annealing temperature for the reaction conditions is greater than a melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 50 (such as at least 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000) of the non-identical primers and/or the length of the annealing step of the reaction conditions is greater than 5 minutes (such as at least 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes; and wherein at least 50 (such as at least 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000) non-identical target human loci are simultaneously amplified; and (iii) detecting the amplified products such as by sequencing the amplified products or hybridizing the amplified products to an array. In some embodiments, the method includes empirically measuring or calculating (such as calculating with a computer) the melting temperature of at least 25, 50, 80, 90, 92, 94, 96, 98, 99, or 100% of the primers in the library and selecting an annealing temperature that satisfies any of these embodiments for PCR amplification of target loci.

In some embodiments, the method involves (i) contacting a sample comprising target human loci with a library of at least 2 (such as at least 5, 10, 25 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000) non-immobilized, non-identical primers that simultaneously hybridize to at least 2 (such as at least 5, 10, 25 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500;  $10,000;\ 15,000;\ 19,000;\ 20,000;\ 25,000;\ 27,000;\ 28,000;$ 30,000; 40,000; 50,000; 75,000; 100,000) non-identical target human loci to produce a reaction mixture; wherein the primers do not include molecular inversion probes (MIPs); (ii) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products comprising target amplicons; wherein at least 2 (such as at least 5, 10, 25 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000;

## US 11,519,035 B2

4

7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000) nonidentical target human loci are simultaneously amplified; and (iii) detecting the amplified products such as by sequencing the amplified products or hybridizing the amplified products to an array. In various embodiments, (i) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 2, 5, 10, 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 10 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers; (ii) the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15 15° C., inclusive) greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 2, 5, 10, 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of 20 the non-identical primers; (iii) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the highest melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers; (iv) the annealing temperature is between 1 and 15° C. (such as 25 between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the highest melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers; (v) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 30 or 15° C. greater than the average melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 2, 5, 10, 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of 35 the non-identical primers; or (vi) the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the average melting temperature (such as the empirically measured or calculated  $T_m$ ) of at 40 least 2, 5, 10, 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers. In various embodiments, (i) the length of the annealing step (per PCR cycle) is greater 45 than 1, 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes or (ii) the length of the annealing step (per PCR cycle) is between 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive. In various embodiments, any of the embodiments for annealing tem- 50 perature are combined with any of the embodiments for annealing time. In various embodiments, the annealing temperature is at least 3° C. greater than the melting temperature of at least 50 of the non-identical primers, the annealing temperature is at least 3° C. greater than the 55 highest melting temperature of the primers, the annealing temperature is at least 8° C. greater than the highest melting temperature of the primers, the annealing temperature is at least 3° C. greater than the average melting temperature of the primers, the annealing temperature is at least 8° C. 60 greater than the average melting temperature of the primers, the range of melting temperature of the primers is between 1 to 5° C., inclusive, the range of melting temperatures of the primers is less than 5° C., or any combination thereof. In some embodiments, the method includes empirically mea- 65 suring or calculating (such as calculating with a computer)

the melting temperature of at least 25, 50, 80, 90, 92, 94, 96,

6

98, 99, or 100% of the primers in the library and selecting an annealing temperature that satisfies any of these embodiments for PCR amplification of target loci. In some embodiments, a crowding agent, such as PEG or glycerol is included in the reaction mixture.

In various embodiments of any of the aspects of the invention, the method includes non-specifically amplifying nucleic acids in a sample comprising target human loci; contacting the amplified nucleic acids with a library of non-identical primers (such as non-immobilized primers) that simultaneously hybridize to at least 1,000 non-identical target human loci to produce a reaction mixture; wherein the primers do not include molecular inversion probes (MIPs); subjecting the reaction mixture to primer extension reaction conditions to produce amplified products comprising target amplicons; wherein at least 1,000 non-identical target human loci are simultaneously amplified; and sequencing the amplified products. In some embodiments, the nonspecific amplification comprises universal polymerase chain reaction (PCR), whole genome application, ligation-mediated PCR, degenerate oligonucleotide primer PCR, or multiple displacement amplification. In some embodiments, the method includes contacting a sample comprising target human loci with a library of non-identical primers (such as non-immobilized primers) that simultaneously hybridize to at least 1,000 non-identical target human loci to produce a reaction mixture; wherein the primers do not include molecular inversion probes (MIPs); subjecting the reaction mixture to primer extension reaction conditions to produce amplified products comprising target amplicons; wherein the annealing temperature for the reaction conditions is greater than the melting temperature of at least 1,000 of the non-identical primers; and wherein at least 1,000 nonidentical target human loci are simultaneously amplified; and sequencing the amplified products. In some embodiments, the method includes contacting a sample comprising target human loci with a library of non-identical primers (such as non-immobilized primers) that simultaneously hybridize to at least 1,000 non-identical target human loci to produce a reaction mixture in which the concentration of each primer is less than 20 nM; wherein the primers do not include molecular inversion probes (MIPs); subjecting the reaction mixture to primer extension reaction conditions to produce amplified products comprising target amplicons; wherein the length of the annealing step of the reaction conditions is greater than 10 minutes; and wherein at least 1,000 non-identical target human loci are simultaneously amplified; and sequencing the amplified products. In some embodiments, the method includes contacting a sample comprising target human loci with a library of non-identical primers (such as non-immobilized primers) that simultaneously hybridize to at least 1,000 non-identical target human loci to produce a reaction mixture; wherein the guaninecytosine (GC) content of the primers is between 30% and 80%, inclusive; wherein the range of melting temperatures of the primers is less than 5° C.; wherein the length of the primers is between 15 to 75 nucleotides, inclusive; and wherein the primers do not include molecular inversion probes (MIPs); subjecting the reaction mixture to primer extension reaction conditions to produce amplified products comprising target amplicons; wherein at least 1,000 nonidentical target human loci are simultaneously amplified; and sequencing the amplified products. In some embodiments, the method does not comprise using a microarray. In some embodiments, the library includes a least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000;

## US 11,519,035 B2

,

40,000; 50,000; 75,000; or 100,000 different primers. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified.

In various embodiments of any of the aspects of the invention, the  $\Delta G$  values for each possible combination of two primers in the library are all equal to or greater than -5 kcal/mol. In some embodiments, the method simultaneously PCR-amplifying at least 1,000 non-identical target human 10 loci in the sample using at least 1,000 non-identical primer pairs (such as non-immobilized primer pairs) to produce a first set of amplified products; wherein each primer pair includes a forward primer and a reverse primer that hybridize to the same target human locus. In some embodiments, 15 the method also includes simultaneously PCR-amplifying at least 1,000 non-identical target human loci in the first set of amplified products using at least 1,000 non-identical primer pairs (such as non-immobilized primer pairs) to produce a second set of amplified products; wherein each primer pair 20 includes a forward primer and a reverse primer that hybridize to the same target human locus. In some embodiments, the primer pairs used in the first and second round of PCR are the same. In some embodiments, the primer pairs used in the first and second round of PCR are different. In some 25 embodiments, the forward primers used in the first and second round of PCR are the same, and the reverse primers used in the first and second round of PCR are different. In some embodiments, the forward primers used in the first and second round of PCR are different, and the reverse primers 30 used in the first and second round of PCR are the same. In some embodiments, the method simultaneously PCR-amplifying at least 1,000 non-identical target human loci in the sample using at least 1,000 non-identical primer pairs (such as non-immobilized primer pairs) to produce a first set of 3: amplified products; wherein each primer pair includes an outer forward primer and an outer reverse primer that hybridize to the same target human locus; and simultaneously PCR-amplifying at least 1,000 non-identical target human loci in the first set of amplified products using a 40 universal reverse primer and at least 1,000 non-identical inner forward primers to produce a second set of amplified products; wherein each inner forward primer hybridizes to a region downstream from the corresponding outer forward primer. In some embodiments, the method includes simul- 45 taneously PCR-amplifying at least 1,000 non-identical target human loci in the sample using at least 1,000 non-identical primer pairs to produce a first set of amplified products; wherein each primer pair includes an outer forward primer and an outer reverse primer that hybridize to the same target 50 human locus; and simultaneously PCR-amplifying at least 1,000 non-identical target human loci in the first set of amplified products using a universal forward primer and at least 1,000 non-identical inner reverse primers to produce a second set of amplified products; wherein each inner reverse 55 primer hybridizes to a region upstream from the corresponding outer reverse primer. In some embodiments, the method includes simultaneously PCR-amplifying at least 1,000 nonidentical target human loci in the sample using at least 1,000 non-identical forward primers and a universal reverse primer 60 to produce a first set of amplified products; and simultaneously PCR-amplifying at least 1,000 non-identical target human loci in the first set of amplified products using a universal forward primer and at least 1,000 non-identical reverse primers to produce a second set of amplified prod- 65 ucts. In some embodiments, the method includes simultaneously PCR-amplifying at least 1,000 non-identical target

8

human loci in the sample using at least 1,000 non-identical reverse primers and a universal forward primer to produce a first set of amplified products; and simultaneously PCRamplifying at least 1,000 non-identical target human loci in the first set of amplified products using a universal reverse primer and at least 1,000 non-identical forward primers to produce a second set of amplified products. In some embodiments, at least 96% of the primer molecules are extended to form amplified products. In some embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the melting temperature of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers. In some embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the highest melting temperature of the non-identical primers. In some embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the average melting temperature of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers. In some embodiments, the range of melting temperature of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers is between 1 to 5° C., inclusive. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers have 2, 1, or 0 guanines or cytosines in the last 5 bases at the 3' end of the primers. In some embodiments, the sample comprises maternal DNA from the pregnant mother of a fetus and fetal DNA, and wherein the method comprises determining the presence or absence of a fetal chromosome abnormality from the sequencing data. In some embodiments, the sample is from an individual suspected of having cancer or an above normal risk for cancer; and wherein one or more of the target human loci comprises a polymorphism or other mutation associated with an above normal risk for cancer or associated with cancer.

In various embodiments of any of the aspects of the invention, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products are target amplicons. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In various embodiments, less than 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.05% of the amplified products are primer dimers. In some embodiments, the library of test primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 test primer pairs, wherein each pair of primers

## US 11,519,035 B2

includes a forward test primer and a reverse test primer that hybridize to the same target locus. In some embodiments, the library of test primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 individual test primers that hybridize to different target loci, wherein the individual primers are not part of primer pairs.

In various embodiments of any of the aspects of the invention, the concentration of each test primer is less than 10 100, 75, 50, 25, 10, 5, 2, 1, 0.5, 0.1, or 0.05 nM. In various embodiments, the guanine-cytosine (GC) content of the test primers is between 30 to 80%, such as between 20 to 70%, 40 to 70%, or 50 to 60%, inclusive. In some embodiments, the range of GC content (e.g., the maximum GC content 15 minus minimum GC content, such as 80%-60%=a range of 20%) of the test primers is less than 30, 20, 10, or 5%. In some embodiments, there are at least 2 (such as 3, 4, or 5) guanines or cytosines in the last 5 bases at the 3' end of the primers. In some embodiments, a maximum of 2 (such as 2, 20 1, or 0) of the bases in the last 5 bases at the 3' end of the primers are guanines or cytosines. In some embodiments, there are at least 1 (such as 2 or 3) guanines or cytosines in the last 3 bases at the 3' end of the primers. In some embodiments, the melting temperature  $(T_m)$  of the test 25 primers is between 40 to 80° C., such as 50 to 70° C., 55 to 65° C., 54 to 60.5° C., or 57 to 60.5° C., inclusive. In some embodiments, the range of melting temperatures of the test primers is less than 20, 15, 10, 5, 3, or 1° C. In some embodiments, the length of the test primers is between 15 to 30 100 nucleotides, such as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, or 20 to 65 nucleotides, inclusive. In some embodiments, the test primers include a tag that is not target specific, such as a tag that forms an internal loop structure. In some embodi- 3: ments, the tag is between two DNA binding regions. In various embodiments, the test primers include a 5' region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In various 40 embodiments, the length of the 3' region is at least 7 nucleotides. In some embodiments, the length of the 3' region is between 7 and 20 nucleotides, such as between 7 to 15 nucleotides, or 7 to 10 nucleotides, inclusive. In various embodiments, the test primers include a 5' region 45 that is not specific for a target locus (such as a tag or a universal primer binding site) followed by a region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In some 50 embodiments, the range of the length of the test primers is less than 50, 40, 30, 20, 10, or 5 nucleotides. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodi- 55 ments, the length of the target amplicons is at least 100; 200; 300; 400; 500; 600; 700; 800; 900; 1,000; 1,200; 1,500; 2,000; or 3,000 nucleotides. In some embodiments, the length of the target amplicons is between 100 and 1,500 nucleotides, such as between 100 to 1,000; 100 to 500, 500 60 to 750, or 750 to 1,000 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 65 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 or all of the target amplicons have a length that falls

10

within the range of the average length of the amplicons±5% of the average length, average length±20%, average length±20%, average length±50%.

In various embodiments of any of the aspects of the invention, the primer extension reaction conditions are polymerase chain reaction conditions (PCR). In various embodiments, the length of the annealing step is greater than 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes. In various embodiments, the length of the extension step is greater than 0.2, 0.5, 1, 3, 5, 8, 10, or 15 minutes.

In various embodiments of any of the aspects of the invention, the test primers are used to simultaneously amplify at least 300 different target loci in a sample that includes maternal DNA from the pregnant mother of a fetus and fetal DNA to determine the presence or absence of a fetal chromosome abnormality. In various embodiments, the method includes ligating a universal primer binding site to the DNA molecules in the sample; amplifying the ligated DNA molecules using at least 300 specific primers and a universal primer to produce a first set of amplified products; and amplifying the first set of amplified products using at least 300 pairs of specific primers to produce a second set of amplified products.

In various embodiments of any of the aspects of the invention, the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in sample includes DNA from an alleged father of a fetus and to simultaneously amplify the target loci in sample that includes maternal DNA from the pregnant mother of the fetus and fetal DNA to establish whether the alleged father is the biological father of the fetus.

In various embodiments of any of the aspects of the invention, the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in one cell or multiple cells from an embryo to determine the presence or absence of a chromosome abnormality. In various embodiments, cells from a set of two or more embryos are analyzed, and one embryo is selected for in vitro fertilization.

In various embodiments of any of the aspects of the invention, the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in a forensic nucleic acid sample. In various embodiments, the length of the annealing step is greater than 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes

In various embodiments of any of the aspects of the invention, the method involves using the test primers to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in a control nucleic acid sample to produce a first set of target amplicons and to simultaneously amplify the target loci in a test nucleic acid sample to produce a second set of target amplicons; and comparing the first and second sets of target amplicons to determine whether a target locus is present in one sample but absent in the other, or whether a target locus is present at different levels in the control sample and the test sample. In various embodiments, the test sample is from an individual suspected of having a disease or phenotype of interest (such as cancer), or an increased risk (such as an above normal level of risk) for a disease or phenotype of interest; and wherein one or more of the target loci include a sequence (e.g., a polymorphism or other mutation) associated with an increased risk (such as an above normal level of risk) for the disease or phenotype of interest, or associated with the disease or phenotype of interest. In various embodiments, the method involves using the test primers to simultaneously

## US 11,519,035 B2

1

amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in a control sample that includes RNA to produce a first set of target amplicons and to simultaneously amplify the target loci in a test sample that includes RNA to produce a second set of target amplicons; and comparing the first and second sets of target amplicons to determine the presence or absence of a difference in the RNA expression levels between the control sample and the test sample. In various embodiments, the RNA is mRNA. In various embodiments, the test sample is from an individual suspected of having a 10 disease or phenotype of interest (such as cancer) or an increased risk for the disease or phenotype of interest (such as cancer); and wherein one or more of the target loci includes a sequence (e.g., a polymorphism or other mutation) associated with an increased risk for the disease or 15 phenotype of interest or associated with the disease or phenotype of interest. In some embodiments, the test sample is from an individual diagnosed with a disease or phenotype of interest (such as cancer); and wherein a difference in the RNA expression level between the control sample and test 20 sample indicates a target locus includes a sequence (e.g., a polymorphism or other mutation) associated with an increased or decreased risk for the disease or phenotype of

In some embodiments of any of the aspects of the invention, the test primers are selected from a library of candidate primers based on one or more parameters, such as the selection of primers using any of the methods of the invention. In some embodiments, the test primers are selected from a library of candidate primers based at least in part on 30 the ability of the candidate primers to form primer dimers.

In one aspect, the invention features methods of selecting test primers from a library of candidate primers. In various embodiments, the selection involves (i) calculating on a computer a score (such as an undesirability score) for most 3: or all of the possible combinations of two candidate primers from the library, wherein each score (such as an undesirability score) is based at least in part on the likelihood of dimer formation between the two candidate primers; (ii) removing the candidate primer with the highest score (such 40 as an undesirability score) from the library of candidate primers; and (iii) if the candidate primer removed in step (ii) is a member of a primer pair, then removing the other member of the primer pair from the library of candidate primers; and (iv) optionally repeating steps (ii) and (iii), 45 thereby selecting a library of test primers. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below a minimum threshold. In some embodiments, the 50 selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number. In various embodiments, a score (such as an undesirability score) is calculated for at least 80, 90, 95, 98, 99, or 99.5% of the possible combinations of candidate primers 55 in the library. In various embodiments, the candidate primers remaining in the library are capable of simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 60 100,000 different target loci. In various embodiments, the method also includes (v) contacting a nucleic acid sample that includes target loci with the candidate primers remaining in the library to produce a reaction mixture; and (vi) subjecting the reaction mixture to primer extension reaction 65 conditions to produce amplified products that includes target amplicons.

12

In one aspect, the invention features methods of selecting test primers from a library of candidate primers. In various embodiments, the selection of test primers are selected from a library of candidate primers involves (i) calculating on a computer a score (such as an undesirability score) for most or all of the possible combinations of two candidate primers from the library, wherein each score (such as an undesirability score) is based at least in part on the likelihood of dimer formation between the two candidate primers; (ii) removing from the library of candidate primers the candidate primer that is part of the greatest number of combinations of two candidate primers with a score (such as an undesirability score) above a first minimum threshold; (iii) if the candidate primer removed in step (ii) is a member of a primer pair, then removing the other member of the primer pair from the library of candidate primers; and (iv) optionally repeating steps (ii) and (iii), thereby selecting a library of test primers. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the first minimum threshold. In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number. In various embodiments, a score (such as an undesirability score) is calculated for at least 80, 90, 95, 98, 99, or 99.5% of the possible combinations of candidate primers in the library. In various embodiments, the candidate primers remaining in the library are capable of simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci. In various embodiments, the method also includes (v) contacting a nucleic acid sample that includes target loci with the candidate primers remaining in the library to produce a reaction mixture; and (vi) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that includes target amplicons.

In various embodiments of any of the aspects of the invention, the selection method involves further reducing the number of candidate primers remaining in the library by decreasing the first minimum threshold used in step (ii) to a lower second minimum threshold and optionally repeating steps (ii) and (iii). In some embodiments, the selection method involves increasing the first minimum threshold used in step (ii) to a higher second minimum threshold and optionally repeating steps (ii) and (iii). In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the second minimum threshold, or until the number of candidate primers remaining in the library is reduced to a desired number.

In various embodiments of any of the aspects of the invention, the method involves, prior to step (i), identifying or selecting primers that hybridize to the target loci. In some embodiments, multiple primers (or primer pairs) hybridize to the same target locus, and the selection method is used to select a one primer (or one primer pair) for this target locus based on one or more parameters. In various embodiments, the method involves, prior to step (ii), removing a primer pair from the library that produces a target amplicon that overlaps with a target amplicon produced by another primer pair. In various embodiments, a candidate primer is selected out of a group of two or more candidate primers with equal scores (such as undesirability scores) for removal from the library of candidate primers based on one or more other

## US 11,519,035 B2

1

parameters. In some embodiments, the candidate primers remaining in the library are used as a library of test primers in any of the methods of the invention. In some embodiments, the resulting library of test primers includes any of the primer libraries of the invention.

In various embodiments of any of the aspects of the invention, the selection method selects candidate primers and divides them into different pools (e.g., 2, 3, 4, 5, 6, or more different pools). Each pool can be used to simultaneously amplify a large number of target loci (or a subset of 10 target loci) in a single reaction volume. In some embodiments, a graph coloring algorithm is used to divide candidate primers into different pools. If desired, this method can be used to minimize the number of different pools needed to amplify most or all of the target loci.

In some embodiments, most or all of the target loci (such as at least 70, 80, 90, 92, 94, 96, 98, 99, or 100% of the target loci) are amplified by at least 2, 3, 4, 5, 6, or more different pools. In some embodiments, most or all of the bases in the target loci (such as at least 70, 80, 90, 92, 94, 96, 98, 99, or 20 100% of the bases in the target loci) are amplified by at least 2,3, 4, 5, 6, or more different pools. In some embodiments, most or all of the bases in the target loci (such as at least 70, 80, 90, 92, 94, 96, 98, 99, or 100% of the bases in the target loci) are amplified by at least 2, 3, 4, 5, 6, or more different 25 primers or primer pairs in different pools. For example, a particular base in a target locus may be amplified by at least 2, 3, 4, 5, 6, or more different primers or primer pairs; wherein each different primer or primer pair is in a different pool. Using different primers or primer pairs to amplify each 30 base allows multiple independent measurements of the base to be made, thereby increasing the accuracy of the method. Dividing the different primers or primer pairs that amplify the same base into different pools prevents interference due to overlapping amplicons being formed by different primers 35 or primer pairs.

In one aspect, the invention features methods of selecting test primers from a library of candidate primers to form 2 or more different primer pools. In various embodiments, the selection involves (i) calculating on a computer a score 40 (such as an undesirability score) for most or all of the possible combinations of two candidate primers from the library, wherein each score (such as an undesirability score) is based at least in part on the likelihood of dimer formation between the two candidate primers; (ii) removing the can- 45 didate primer with the highest score (such as an undesirability score) from the library of candidate primers; and (iii) if the candidate primer removed in step (ii) is a member of a primer pair, then removing the other member of the primer pair from the library of candidate primers; and (iv) option- 50 ally repeating steps (ii) and (iii), thereby selecting a first pool. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below a minimum threshold for the first 55 pool. In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number for the first pool. In some embodiments, after the first pool is selected those primers are removed from further consideration and steps of 60 the method (such as steps (ii) and (iii)) are repeated with the remaining primers to select a second pool. If desired, this method may be repeated to select the desired number of primer pools. In various embodiments, a score (such as an undesirability score) is calculated for at least 80, 90, 95, 98, 65 99, or 99.5% of the possible combinations of candidate primers in the library. In some embodiments, the score is

14

based at least in part on the current coverage of the bases in the target locus (such as the number of other primer pools that have a primer or primer pair that amplifies a particular base in the target locus). In various embodiments, one or more of the pools are each capable of simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci. In various embodiments, the method also includes separately contacting a nucleic acid sample that includes target loci with two or more of the pools to produce separate reaction mixtures; and (vi) subjecting the reaction mixtures to primer extension reaction conditions to produce amplified products that includes target 15 amplicons.

In one aspect, the invention features methods of selecting test primers from a library of candidate primers to form 2 or more different primer pools. In various embodiments, the selection of test primers are selected from a library of candidate primers involves (i) calculating on a computer a score (such as an undesirability score) for most or all of the possible combinations of two candidate primers from the library, wherein each score (such as an undesirability score) is based at least in part on the likelihood of dimer formation between the two candidate primers; (ii) removing from the library of candidate primers the candidate primer that is part of the greatest number of combinations of two candidate primers with a score (such as an undesirability score) above a first minimum threshold; (iii) if the candidate primer removed in step (ii) is a member of a primer pair, then removing the other member of the primer pair from the library of candidate primers; and (iv) optionally repeating steps (ii) and (iii), thereby selecting a first pool. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the first minimum threshold for the first pool. In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number for the first pool. In various embodiments, the selection method involves further reducing the number of candidate primers remaining in the library by decreasing the first minimum threshold used in step (ii) to a lower second minimum threshold and optionally repeating steps (ii) and (iii). In some embodiments, the selection method involves increasing the first minimum threshold used in step (ii) to a higher second minimum threshold and optionally repeating steps (ii) and (iii). In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the second minimum threshold, or until the number of candidate primers remaining in the library is reduced to a desired number for the first pool. In some embodiments, after the first pool is selected those primers are removed from further consideration and steps of the method (such as steps (ii) and (iii)) are repeated with the remaining primers to select a second pool. If desired, this method may be repeated to select the desired number of primer pools. In various embodiments, a score (such as an undesirability score) is calculated for at least 80, 90, 95, 98, 99, or 99.5% of the possible combinations of candidate primers in the library. In some embodiments, the score is based at least in part on the current coverage of the bases in the target locus (such as the number of other primer pools that have a primer or primer pair that amplifies a particular base in the target locus). In various embodiments, one or more of the pools are

## US 11,519,035 B2

1:

each capable of simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci. In various embodiments, the method also includes separately contacting a nucleic acid sample that includes target loci with two or more of the pools to produce separate reaction mixtures; and (vi) subjecting the reaction mixtures to primer extension reaction conditions to produce amplified products that includes target amplicons.

In some embodiments, at least 70, 80, 85, 90, 95, or 100% of the nucleotides in a region of interest (such as an exon) are included in at least 1, 2, 3, or 4 different amplicons (i.e., amplicons with non-identical sequences that are formed by different primers or primer pairs). In some embodiments, at 15 least 70, 80, 85, 90, 95, or 100% of the nucleotides in at least 70, 80, 85, 90, 95, or 100% of the regions of interest are amplified by at least 1, 2, 3, or 4 different amplicons. In some embodiments in which 2-fold coverage is desired, the primers are divided into at least two different pools such the 20 amplicons in each pool do not overlap with each other (which would cause interference during amplification).

In various embodiments of any of the aspects of the invention, the score (such as the undesirability score) are based at least in part on one or more parameters selected 25 from the group consisting of heterozygosity rate of the target locus, disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, specificity of the candidate primer for the target locus, size of the candidate primer, melting temperature of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target 35 amplicon, and size of the target amplicon.

In various embodiments of any of the aspects of the invention, the score (such as the undesirability score) are based at least in part on one or more parameters selected from the group consisting of heterozygosity rate of the target 40 locus, specificity of the candidate primer for the target locus; size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the 45 candidate primer, amplification efficiency of the target amplicon, and size of the target amplicon; and the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in a sample that includes maternal DNA from the pregnant mother of a 50 fetus and fetal DNA to determine the presence or absence of a fetal chromosome abnormality. In various embodiments, the method includes ligating a universal primer binding site to the DNA molecules in the sample; amplifying the ligated DNA molecules using e.g. at least 100 (e.g., at least 300 or 55 1,000) specific primers and a universal primer to produce a first set of amplified products; and amplifying the first set of amplified products using e.g. at least 100 (e.g., at least 300 or 1,000) pairs of specific primers to produce a second set of amplified products. In various embodiments, at least 500; 60 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primer pairs are used. In various embodiments, at least 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 65 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75;

16

100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold.

In various embodiments of any of the aspects of the invention, the score (such as the undesirability score) are based at least in part on one or more parameters selected from the group consisting of heterozygosity rate of the target locus, specificity of the candidate primer for the target locus; size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target amplicon, and size of the target amplicon; and the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in sample includes DNA from an alleged father of a fetus and to simultaneously amplify the target loci in a sample that includes maternal DNA from the pregnant mother of a fetus and fetal DNA to establish whether the alleged father is the biological father of the fetus. In various embodiments, at least 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold.

In various embodiments of any of the aspects of the invention, the score (such as the undesirability score) are based at least in part on one or more parameters selected from the group consisting of heterozygosity rate of the target locus, specificity of the candidate primer for the target locus; size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target amplicon, and size of the target amplicon; and the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in one cell or multiple cells from an embryo to determine the presence or absence of a chromosome abnormality. In various embodiments, cells from a set of two or more embryos are analyzed, and one embryo is selected for in vitro fertilization. In various embodiments, at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold.

## US 11,519,035 B2

17

In various embodiments of any of the aspects of the invention, the scores (such as the undesirability scores) are based at least in part on one or more parameters selected from the group consisting of heterozygosity rate of the target locus, specificity of the candidate primer for the target locus; 5 size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target 10 amplicon, and size of the target amplicon; and the test primers are used to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in a forensic nucleic acid sample. In various embodiments, the length of the annealing step is greater than 3, 5, 8, 10, 15, 20, 30, 45, 15 60, 75, 90, 120, 150, or 180 minutes. In various embodiments, at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 20 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 25 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold.

In various embodiments of any of the aspects of the invention, the scores (such as the undesirability scores) are 30 based at least in part on one or more parameters selected from the group consisting of heterozygosity rate of the target locus, disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, disease penetrance associated with a sequence (e.g., a polymorphism) at the target 3 locus, specificity of the candidate primer for the target locus, size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the 40 candidate primer, amplification efficiency of the target amplicon, and size of the target amplicon; and the method involves using the test primers to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000) different target loci in a control nucleic acid sample to produce a first set of 45 target amplicons and to simultaneously amplify the target loci in a test nucleic acid sample to produce a second set of target amplicons; and comparing the first and second sets of target amplicons to determine whether a target locus is present in one sample but absent in the other, or whether a 50 target locus is present at different levels in the control sample and the test sample. In various embodiments, the test sample is from an individual suspected of having a disease or phenotype of interest, or an increased risk for a disease or phenotype of interest; and wherein one or more of the target 55 loci include a sequence (e.g., a polymorphism) at the target locus associated with an increased risk for the disease or phenotype of interest, or associated with the disease or phenotype of interest. In various embodiments, at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 60 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 65 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or

18

400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold.

In various embodiments of any of the aspects of the invention, the scores (such as the undesirability scores) are based at least in part on one or more parameters selected from the group consisting of heterozygosity rate of the target locus, disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, specificity of the candidate primer for the target locus, size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target amplicon, and size of the target amplicon; and the method involves using the test primers to simultaneously amplify e.g. at least 100 (e.g., at least 300 or 1,000); different target loci in a control sample that includes RNA to produce a first set of target amplicons and to simultaneously amplify the target loci in a test sample that includes RNA to produce a second set of target amplicons; and comparing the first and second sets of target amplicons to determine the presence or absence of a difference in the RNA expression levels between the control sample and the test sample. In various embodiments, the RNA is mRNA.

In various embodiments, the test sample is from an individual suspected of having a disease or phenotype of interest (such as cancer) or an increased risk for the disease or phenotype of interest (such as cancer); and wherein one or more of the target loci includes a sequence (e.g., a polymorphism or other mutation) associated with an increased risk for the disease or phenotype of interest or associated with the disease or phenotype of interest. In some embodiments, the test sample is from an individual diagnosed with a disease or phenotype of interest (such as cancer); and wherein a difference in the RNA expression level between the control sample and test sample indicates a target locus includes a sequence (e.g., a polymorphism or other mutation) associated with an increased or decreased risk for the disease or phenotype of interest. In various embodiments, at least 300, 500; 750; 1,000; 2000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold.

In one aspect, the invention features libraries of primers (such as non-immobilized primers). In some embodiments, the primers are selected from a library of candidate primers using any of the methods of the invention. In some embodiments, the library includes primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci. In some embodiments, the library includes primers that simultaneously amplify at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000;

## US 11,519,035 B2

10

30,000; 40,000; 50,000; 75,000; or 100,000 different target loci. In some embodiments, the library includes primers that simultaneously amplify at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci such that less than 60, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.05% of the amplified products are primer dimers. In some embodiments, the library includes primers that simultaneously amplify at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 10 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci such that at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products are target amplicons. In some embodiments, the library includes prim- 15 ers that simultaneously amplify target loci such that at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci out of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 20 100,000 different target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 25 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, the library of primers includes at least 30 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500;  $10,000;\ 15,000;\ 19,000;\ 20,000;\ 25,000;\ 27,000;\ 28,000;$ 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers or primer pairs. In some embodiments, the library of primers includes at least 25; 50; 75; 100; 300; 500; 750; 35 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 primer pairs, wherein each pair of primers includes a forward test primer and a reverse test primer where each pair of test primers hybridize to a target locus. In some 40 embodiments, the library of primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 individual primers that each hybridize to a different target locus, wherein the 45 individual primers are not part of primer pairs. In some embodiments, the primers in the library are not immobilized (such as not immobilized to a solid support) or not part of a microarray. In some embodiments, the primers are dissolved in solution (such as dissolved in the liquid phase). In some 50 embodiments, the library of primers consists essentially of, or consists of primers.

In some embodiments,  $\Delta G$  values for each possible combination of two primers (each possible primer dimer) in a library are all equal to or greater than -20, -18, -16, -14, 55, -12, -10, -9, -8, -7, -6, -5, -4, -3, -2, or -1 kcal/mol. In some embodiments,  $\Delta G$  values for at least 80, 85, 90, 92, 94, 96, 98, 99, or 100% of the primers in the library for possible combinations of that primer with other primers in the library are all equal to or greater than -20, -18, -16, -14, -12, -10, 60, -9, -8, -7, -6, -5, -4, -3, -2, or -1 kcal/mol.

In various embodiments of any of the aspects of the invention, the concentration of each primer is less than 100, 75, 50, 25, 10, 5, 2, 1, 0.5, 0.1, or 0.05 nM. In various embodiments, the GC content of the primers is between 30 65 to 80%, such as between 40 to 70%, 20 to 70%, or 50 to 60%, inclusive. In some embodiments, the range of GC

20

content of the primers is less than 30, 20, 10, or 5%. In some embodiments, there are at least 2 (such as 3, 4, or 5) guanines or cytosines in the last 5 bases at the 3' end of the primers. In some embodiments, there are at least 1 (such as 2 or 3) guanines or cytosines in the last 3 bases at the 3' end of the primers. In some embodiments, a maximum of 2 (such as 2, 1, or 0) bases in the last 5 bases at the 3' end of the primers are guanines or cytosines. In some embodiments, the melting temperature of the primers is between 40 to 80° C., such as 50 to 70° C., 55 to 65° C., 54 to 60.5° C., or 57 to 60.5° C., inclusive. In some embodiments, the range of melting temperature of the primers is less than 15, 10, 5, 3, or 1° C. In some embodiments, the length of the primers is between 15 to 100 nucleotides, such as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, or 20 to 65 nucleotides, inclusive. In some embodiments, the primers include a tag that is not target specific, such as a tag that forms an internal loop structure. In some embodiments, the tag is between two DNA binding regions. In various embodiments, the primers include a 5' region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In various embodiments, the length of the 3' region is at least 7 nucleotides. In some embodiments, the length of the 3' region is between 7 and 20 nucleotides, such as between 7 to 15 nucleotides, or 7 to 10 nucleotides, inclusive. In various embodiments, the primers include a 5' region that is not specific for a target locus (such as another tag or a universal primer binding site) followed by a region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In some embodiments, the range of the length of the primers is less than 50, 40, 30, 20, 10, or 5 nucleotides. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the target loci are on two or more different chromosomes, such as two or more of chromosomes 13, 18, 21, X and Y. In some embodiments, the target loci are target human loci. In some embodiments, the target loci includes a sequence (e.g., a polymorphism or other mutation) associated with an increased risk for the disease or phenotype of interest (such as cancer), or associated with the disease or phenotype of interest (such as cancer). In some embodiments, the polymorphism or mutation is a driver mutation that has a causative role in the disease or phenotype of interest (such as cancer). In some embodiments, the polymorphism or mutation is not a causative mutation. For example, in some cancers, multiple mutations accumulate but some of them are not causative mutations. Polymorphisms or mutations (such as those that are present at a higher frequency in subjects with a disease or phenotype of interest such as cancer than subjects without the disease or phenotype of interest such as cancer) that are not causative can still be useful for diagnosing the disease or phenotype. In some embodiments, the polymorphisms or mutation is present at a higher frequency in subjects with a disease or disorder (such as cancer) than subjects (such as healthy or normal subjects) without the disease or disorder (such as cancer). In some embodiments, the polymorphisms or mutation is indicative of cancer, such as a causative mutation. In some embodiments, the polymorphism(s) or mutation(s) are directly detected. In some embodiments,

## US 11,519,035 B2

2

polymorphism(s) or mutation(s) are indirectly detected by detection of one or more sequences (e.g., a polymorphic locus such as a SNP) that are linked to the polymorphism or mutation).

In one aspect, the invention provides a composition that 5 includes any of the primer libraries of the invention (such as non-immobilized primers). In some embodiments, the composition includes one or more free nucleotides (such as deoxynucleotides, ATP, CTP, GTP, TTP, UTP, dATP, dCTP, dGTP, dTTP, dUTP an activated nucleotide or deoxynucle- 10 otide, or a non-naturally occurring nucleotide or deoxynucleotide). In some embodiments, the composition includes at least one primer (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 15 50,000; 75,000; or 100,000 non-identical primers) with a polynucleotide sequence of a human nucleic acid and at least one primer (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 20 100,000 non-identical primers) with a polynucleotide sequence that is not found in a human (such as a universal primer, a primer that comprises a region or consists entirely of random nucleotides, or a primer with a region such as a tag or barcode of one or more nucleotides that are not found 25 in a human or are not found in nature as part of the polynucleotide sequence of the primer). In some embodiments, the composition includes at least one primer (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 30 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 nonidentical primers) with the polynucleotide sequence of a human nucleic acid and one or more non-human or nonnaturally occurring enzymes (e.g., ligase or polymerase from a species other than a human, such as a bacterial or 35 non-naturally-occurring ligase or polymerase). In some embodiments, the composition includes at least one primer (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 40 non-identical primers) with the polynucleotide sequence of a human nucleic acid and a buffer or additive that is non-naturally-occurring or is not found in a human. In some embodiments, the composition comprises, consists essentially of, or consists of one or more of the following: 45 primer(s), amplicon(s) free nucleotide(s), non-human or non-naturally occurring enzyme(s), buffer(s), additive(s), or any combination thereof. In some embodiments, the composition comprises, consists essentially of, or consists of primers and one or more non-human or non-naturally occur- 50 ring enzymes. Exemplary non-naturally occurring enzymes contain at least one sequence difference compared to naturally occurring (wild-type) enzymes.

In one aspect, the invention provides a composition comprising at least 100 different amplicons (e.g., at least 55 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical amplicons) in solution in one reaction volume. In some embodiments, the amplicons are produced from the simultaneous PCR amplification of at least 100 different target loci (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical target loci) using at least 100 different primers or primer 65 pairs (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000;

22

25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers or primer pairs) in one reaction volume. In some embodiments, (i) less than 60% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (ii) less than 40% of the amplified products are primer dimers and at least 60% of the amplified products are target amplicons, (iii) less than 20% of the amplified products are primer dimers and at least 80% of the amplified products are target amplicons, (iv) less than 10% of the amplified products are primer dimers and at least 90% of the amplified products are target amplicons, or (v) less than 5% of the amplified products are primer dimers and at least 95% of the amplified products are target amplicons. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 and 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the range of the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, the composition includes at least 1,000 different amplicons in solution in one reaction volume. In some embodiments, the amplicons are produced from the simultaneous PCR amplification of at least 1,000 different target human loci using at least 1,000 different primers in one reaction volume; wherein (i) less than 20% of the amplicons are primer dimers, and (ii) at least 80% of the amplicons comprise one of the target human loci and are between 50 and 100 nucleotides in length, inclusive. In some embodiments, the composition consists essentially of, or consists of one or more of the following: amplicons, primers (such as any of the primers disclosed herein), free nucleotide(s), non-human or non-naturally occurring enzyme(s), buffer(s), or any combination thereof. In some embodiments, at least one amplicon or primer has a nonhuman or non-naturally occurring sequence, nucleotide, or linkage between nucleotides.

In one aspect, the invention provides a composition comprising at least 100 different primers or primer pairs (e.g., at least 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers or primer pairs) and at least 100 different amplicons (e.g., at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical amplicons) in solution in one reaction volume. In some embodiments, the amplicons are produced from the simultaneous PCR amplification of at least 100 different target loci (e.g., at least 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical target loci) using the primers or primer pairs in one reaction volume. In some embodiments, (i) less than 60% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (ii) less than 40% of the amplified products are primer dimers and at least 60% of the amplified products are target amplicons, (iii) less than 20% of the amplified products are primer dimers and at least 80% of the amplified products are target amplicons, (iv) less than 10% of the amplified products are primer dimers and at least 90% of the amplified products are target amplicons, or (v) less than 5% of the amplified products are primer dimers and at least 95% of the amplified products are target amplicons. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60

## US 11,519,035 B2

2

and 80 nucleotides, or 60 and 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the range of the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 5 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, the composition comprising at least 1,000 different primers and at least 1,000 different amplicons in solution in one reaction volume. In some embodiments, the amplicons are produced from the 10 simultaneous PCR amplification of at least 1,000 different target human loci with the primers in one reaction volume; wherein (i) less than 20% of the amplicons are primer dimers, and (ii) at least 80% of the amplicons comprise one of the target loci and are between 50 and 100 nucleotides in 15 length, inclusive. In some embodiments, the composition consists essentially of, or consists of one or more of the following: amplicons, primers (such as any of the primers disclosed herein), free nucleotide(s), non-human or nonnaturally occurring enzyme(s), buffer(s), or any combination 20 thereof. In some embodiments, at least one amplicon or primer has a non-human or non-naturally occurring sequence, nucleotide, or linkage between nucleotides.

In one aspect, the invention provides kits that include any of the primer libraries or compositions of the invention for 25 amplifying target loci in a nucleic acid sample. In some embodiments, the kits consists essentially of, or consists of primers, primers and instructions for using the primers, a composition of the invention, or a composition of the invention and instructions for using the composition. In 30 some embodiments, the kit includes instructions for using the library to amplify the target loci.

In one aspect, the invention provides an apparatus, device, or composition that includes any of the primer libraries or compositions of the invention. In some embodiments, the 35 apparatus, device, or composition includes a physical structure (such as one or more reaction vessels, reaction chambers, or wells) that contains the primer library or composition of the invention (for example, the primers may be dissolved in a solution that is in the physical structure). In some embodiments, the physical structure is a non-naturally occurring physical structure or a physical structure that does not naturally contain a primer library or composition of the invention (such as a physical structure that is not found in nature with nucleic acids in it).

In one aspect, the invention features methods for determining a ploidy status of chromosome in a gestating fetus. In some embodiments, the method involves contacting a nucleic acid sample with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 50 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci to produce a reaction mixture; wherein the nucleic acid sample includes maternal DNA from the mother of the fetus and fetal DNA from the 55 fetus. In some embodiments, the reaction mixture is subjected to primer extension reaction conditions to produce amplified products; the amplified products are measured with a high throughput sequencer to produce sequencing data; allele counts at the polymorphic loci are calculated on 60 a computer based on the sequencing data; a plurality of ploidy hypotheses each pertaining to a different possible ploidy state of the chromosome are created on a computer; a joint distribution model for the expected allele counts at the polymorphic loci on the chromosome is built on a 65 computer for each ploidy hypothesis; a relative probability of each of the ploidy hypotheses is determined on a com24

puter using the joint distribution model and the allele counts; and the ploidy state of the fetus is called by selecting the ploidy state corresponding to the hypothesis with the greatest probability.

In one aspect, the invention features methods for determining a ploidy status of a chromosome in a gestating fetus. In an embodiment a method for determining a ploidy status of a chromosome in a gestating fetus includes obtaining a first sample of DNA that comprises maternal DNA from the mother of the fetus and fetal DNA from the fetus, preparing the first sample by isolating the DNA so as to obtain a prepared sample, measuring the DNA in the prepared sample at a plurality of polymorphic loci on the chromosome, calculating, on a computer, allele counts at the plurality of polymorphic loci from the DNA measurements made on the prepared sample, creating, on a computer, a plurality of ploidy hypotheses each pertaining to a different possible ploidy state of the chromosome, building, on a computer, a joint distribution model for the expected allele counts at the plurality of polymorphic loci on the chromosome for each ploidy hypothesis, determining, on a computer, a relative probability of each of the ploidy hypotheses using the joint distribution model and the allele counts measured on the prepared sample, and calling the ploidy state of the fetus by selecting the ploidy state corresponding to the hypothesis with the greatest probability.

In one aspect, the invention features methods of testing for an abnormal distribution of a chromosome in a sample that includes a mixture of maternal and fetal DNA. In some embodiments, the method involves (i) contacting the sample with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci to produce a reaction mixture; wherein the target loci are from a plurality of different chromosomes; and wherein the plurality of different chromosomes include at least one first chromosome suspected of having an abnormal distribution in the sample and at least one second chromosome presumed to be normally distributed in the sample; (ii) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products; (iii) sequencing the amplified products to obtain a plurality of sequence tags aligning to the target loci; wherein the sequence tags are of sufficient length to be assigned to a specific target locus; (iv) assigning on a computer the plurality of sequence tags to their corresponding target loci; (v) determining on a computer a number of sequence tags aligning to the target loci of the first chromosome and a number of sequence tags aligning to the target loci of the second chromosome; and (vi) comparing on a computer the numbers from step (v) to determine the presence or absence of an abnormal distribution of the first chromosome.

In one aspect, the invention provides methods for detecting the presence or absence of a fetal aneuploidy. In some embodiments, the method involves (i) contacting a sample that includes a mixture of maternal and fetal DNA with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different non-polymorphic target loci to produce a reaction mixture; wherein the target loci are from a plurality of different chromosomes; (ii) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that includes target amplicons; (iii) quantifying on a computer a relative frequency of the target amplicons from the

# US 11,519,035 B2

2

first and second chromosomes of interest; (iv) comparing on a computer the relative frequency of the target amplicons from the first and second chromosomes of interest; and (v) identifying the presence or absence of an aneuploidy based on the compared relative frequencies of the first and second chromosome of interest. In some embodiments, the first chromosome is a chromosome suspected of being euploid. In some embodiments, the second chromosome is a chromosome suspected of being aneuploidy.

In one aspect, a method is disclosed for determining presence or absence of fetal aneuploidy in a maternal tissue sample comprising fetal and maternal genomic DNA, the method including (a) obtaining a mixture of fetal and maternal genomic DNA from said maternal tissue sample, (b) conducting massively parallel DNA sequencing of DNA fragments randomly selected from the mixture of fetal and maternal genomic DNA of step (a) to determine the sequence of said DNA fragments, (c) identifying chromosomes to which the sequences obtained in step (b) belong, 20 (d) using the data of step (c) to determine an amount of at least one first chromosome in said mixture of maternal and fetal genomic DNA, wherein said at least one first chromosome is presumed to be euploid in the fetus, (e) using the data of step (c) to determine an amount of a second chro- 25 mosome in said mixture of maternal and fetal genomic DNA, wherein said second chromosome is suspected to be aneuploid in the fetus, (f) calculating the fraction of fetal DNA in the mixture of fetal and maternal DNA, (g) calculating an expected distribution of the amount of the second 30 target chromosome if the second target chromosome is euploid, using the number in step (d), (h) calculating an expected distribution of the amount of the second target chromosome if the second target chromosome is aneuploid, using the first number is step (d) and the calculated fraction 35 of fetal DNA in the mixture of fetal and maternal DNA in step (f), and (i) using a maximum likelihood or maximum a posteriori approach to determine whether the amount of the second chromosome as determined in step (e) is more likely to be part of the distribution calculated in step (g) or the 40 distribution calculated in step (h); thereby indicating the presence or absence of a fetal aneuploidy.

In various embodiments of any of the aspects of the invention, the target loci include one or more SNPs in the homologous non-recombining region of chromosome X and/or chromosome Y. In some embodiments, the method includes determining the relative amount of chromosome X and chromosome Y. In some embodiments, the method includes determining the number of copies of chromosome X and/or chromosome Y.

In some embodiments, the method also includes obtaining genotypic data from one or both parents of the fetus. In some embodiments, obtaining genotypic data from one or both parents of the fetus includes preparing the DNA from the parents where the preparing comprises preferentially enriching the DNA at the plurality of polymorphic loci to give prepared parental DNA, optionally amplifying the prepared parental DNA, and measuring the parental DNA in the prepared sample at the plurality of polymorphic loci.

In various embodiments of any of the aspects of the 60 invention, building a joint distribution model for the expected allele count probabilities of the plurality of polymorphic loci on the chromosome is done using the obtained genetic data from the one or both parents.

In some embodiments, the sample (e.g., the first sample) 65 has been isolated from maternal plasma and where the obtaining genotypic data from the mother is done by esti-

26

mating the maternal genotypic data from the DNA measurements made on the prepared sample.

In one aspect, a diagnostic box is disclosed for helping to determine a ploidy status of a chromosome in a gestating fetus where the diagnostic box is capable of executing the preparing and measuring steps of any of the methods of the invention

In various embodiments of any of the aspects of the invention, the allele counts are probabilistic rather than binary. In some embodiments, measurements of the DNA in the prepared sample at the plurality of polymorphic loci are also used to determine whether or not the fetus has inherited one or a plurality of disease linked haplotypes.

In various embodiments of any of the aspects of the invention, building a joint distribution model for allele count probabilities is done by using data about the probability of chromosomes crossing over at different locations in a chromosome to model dependence between polymorphic alleles on the chromosome. In some embodiments, building a joint distribution model for allele counts and the step of determining the relative probability of each hypothesis are done using a method that does not require the use of a reference chromosome.

In various embodiments of any of the aspects of the invention, determining the relative probability of each hypothesis makes use of an estimated fraction of fetal DNA in the prepared sample. In some embodiments, the DNA measurements from the prepared sample used in calculating allele count probabilities and determining the relative probability of each hypothesis comprise primary genetic data. In some embodiments, selecting the ploidy state corresponding to the hypothesis with the greatest probability is carried out using maximum likelihood estimates or maximum a posteriori estimates.

In various embodiments of any of the aspects of the invention, calling the ploidy state of the fetus also includes combining the relative probabilities of each of the ploidy hypotheses determined using the joint distribution model and the allele count probabilities with relative probabilities of each of the ploidy hypotheses that are calculated using statistical techniques taken from a group consisting of a read count analysis, comparing heterozygosity rates, a statistic that is only available when parental genetic information is used, the probability of normalized genotype signals for certain parent contexts, a statistic that is calculated using an estimated fetal fraction of the sample (e.g., the first sample) or the prepared sample, and combinations thereof.

In various embodiments of any of the aspects of the invention, a confidence estimate is calculated for the called ploidy state. In some embodiments, the method also includes taking a clinical action based on the called ploidy state of the fetus, wherein the clinical action is selected from one of terminating the pregnancy or maintaining the pregnancy.

In various embodiments of any of the aspects of the invention, the method may be performed for fetuses at between 4 and 5 weeks gestation; between 5 and 6 weeks gestation; between 6 and 7 weeks gestation; between 7 and 8 weeks gestation; between 8 and 9 weeks gestation; between 9 and 10 weeks gestation; between 10 and 12 weeks gestation; between 12 and 14 weeks gestation; between 14 and 20 weeks gestation; between 20 and 40 weeks gestation; in the first trimester; in the second trimester; in the third trimester; or combinations thereof.

In various embodiments of any of the aspects of the invention, a report displaying a determined ploidy status of a chromosome in a gestating fetus generated using the method. In some embodiments, a kit is disclosed for deter-

## US 11,519,035 B2

2

mining a ploidy status of a target chromosome in a gestating fetus designed to be used with any of the methods of the invention, the kit including a plurality of inner forward primers and optionally the plurality of inner reverse primers, where each of the primers is designed to hybridize to the 5 region of DNA immediately upstream and/or downstream from one of the polymorphic sites on the target chromosome, and optionally additional chromosomes, where the region of hybridization is separated from the polymorphic site by a small number of bases, where the small number is 10 selected from the group consisting of 1, 2, 3, 4, 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, 31 to 60, and combinations thereof.

In one aspect, the invention features methods for establishing whether an alleged father is the biological father of 15 a fetus that is gestating in a pregnant mother. In some embodiments the method involves, (i) simultaneously amplifying a plurality of polymorphic loci that includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 20 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci on genetic material from the alleged father to produce a first set of amplified products; (ii) simultaneously amplifying the corresponding plurality of polymorphic loci on a mixed sample of DNA originating from a 25 blood sample from the pregnant mother to produce a second set of amplified products; wherein the mixed sample of DNA includes fetal DNA and maternal DNA; (iii) determining on a computer the probability that the alleged father is the biological father of the fetus using genotypic measure- 30 ments based on the first and second sets of amplified products; and (iv) establishing whether the alleged father is the biological father of the fetus using the determined probability that the alleged father is the biological father of the fetus. In various embodiments, the method further 35 includes simultaneously amplifying the corresponding plurality of polymorphic loci on genetic material from the mother to produce a third set of amplified products; wherein the probability that the alleged father is the biological father of the fetus is determined using genotypic measurements 40 based on the first, second, and third sets of amplified products.

In one aspect, the invention provides methods of estimating relative likelihoods that each embryo from a set of embryos will develop as desired. In some embodiments, the 45 method involves contacting a sample from each embryo with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different 50 target loci to produce a reaction mixture for each embryo, wherein the samples are each derived from one or more cells from an embryo. In some embodiments, each reaction mixture is subjected to primer extension reaction conditions to produce amplified products. In some embodiments, the 55 method includes determining on a computer one or more characteristics of at least one cell from each embryo based on the amplified products; and estimating on a computer the relative likelihoods that each embryo will develop as desired, based on the one or more characteristics of the at 60 least one cell for each embryo.

In one aspect, the invention features methods of measuring the amount of two or more target loci in a nucleic acid sample. In some embodiments, the method involves (i) using PCR to amplify a nucleic acid sample that includes a first standard locus, a second standard locus, a first target locus, and a second target locus to form amplified products;

28

wherein the first standard locus and the first target locus have the same number of nucleotides but have a sequence that differs at one or more nucleotides; and wherein the second standard locus and the second target locus have the same number of nucleotides but have a sequence that differs at one or more nucleotides; (ii) sequencing the amplified products to determine a standard ratio that compares the relative amount of the amplified first standard locus compared to the amplified second standard locus; wherein the standard ratio indicates the difference in PCR efficiency for the amplification of the first standard locus and the second standard locus; (iii) determining a target ratio that compares the relative amount of the amplified first target locus compared to the amplified second target locus; and (iv) adjusting the target ratio from step (iii) based on the standard ratio from step (ii) to determine the relative amount of the first target locus and the second target locus in the sample. In various embodiments, the method involves determining the absolute amount of the first target locus and the second target locus in the sample. In various embodiments, the method further includes determining the presence or absence of a target locus (e.g, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci) in the sample. In various embodiments, the method involves using any of the primer libraries of the invention. In various embodiments, the method involves simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci.

In one aspect, the invention features methods of quantitatively measuring a plurality of genetic targets in a sample for analysis. In some embodiments, the method includes (i) mixing genetic material derived from the sample for analysis with a plurality of target specific amplification reagents, and a plurality of standard sequences corresponding to the target specific amplification reagent targets; (ii) amplifying target regions of the genetic material and the standard sequences to produce target amplicons and standard sequence amplicons; and (iii) measuring the quantity of target amplicons and standard sequence amplicons produced. In some embodiments, the genetic material is present in a genetic library. In some embodiments, the genetic targets are polymorphic loci (such as SNPs). In some embodiments, the measuring of quantity is achieved by counting sequences. In some embodiments, the method further includes determining the estimated copy number of at least one chromosome in a sample from which the genetic library was derived, wherein the determination involves comparing the number of sequence reads of a target amplicon with the number of sequence reads of a standard amplicon. In some embodiments, the standard sequences and the genetic library include universal priming sites cable of being primed by the same primer. In some embodiments, the mixing step includes at least 10; 100, 500; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target specific amplification reagents and at least 10; 100, 500; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 standard sequences. In various embodiments, the method involves using any of the primer libraries of the invention. In various embodiments, the method involves simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000;

Case: 24-1324 Document: 42-1 Page: 137 Filed: 03/18/2024

## US 11,519,035 B2

29

75,000; or 100,000 different target regions. In some embodiments, the relative amounts of each of the standard sequences are known. In some embodiments, the relative amounts of each of the sequences is has been calibrated with respect to a reference genome. In some embodiments, the 5 sample for analysis includes a mixture of fetal and maternal genomes. In some embodiments, the sample for analysis is derived from the blood of a pregnant woman or derived from blood plasma. In some embodiments, the reference genome has at least one aneuploidy, such as an aneuploidy at chromosome 13, 18, 21, X, or Y. In some embodiments, the reference genome is diploid.

In one aspect, the invention features a mixture that includes a plurality of genetic standard sequences, wherein 15 the relative amount of each genetic standard sequence in the mixture has been determined by calibration to a reference genome. In various embodiments, the mixture includes at least 10; 100, 500; 1,000; 2,000; 5,000; 7,500; 10,000; 40,000; 50,000; 75,000; or 100,000 genetic standard sequences. In various embodiments, the genetic standard sequences include a first universal priming site, a second universal priming site, a first target specific priming site, a second target specific priming site, and a marker sequence 25 located between the first and second target specific priming sites, wherein the first target specific site and the second target specific priming site are located between the first and second universal priming sites. In various embodiments, the calibration involves using any of the primer libraries of the 30 invention. In various embodiments, the calibration involves simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target regions. In some embodiments, the reference genome has at least one aneuploidy, such as an aneuploidy at chromosome 13, 18, 21, X, or Y. In some embodiments, the reference genome is diploid.

In one aspect, the invention features methods of producing a set of calibrated genetic standard sequences. In some 40 embodiments, the method includes (i) forming an amplification reaction mixture that includes a genetic library prepared from a reference genome, a plurality of target-specific amplification primer reagent sets, and a plurality of genetic standard sequences corresponding to the target specific 45 amplification reagent sets, (ii) amplifying the genetic library and the genetic standard sequences to produce amplicons from the target sequences and amplicons from the genetic standard sequences, (iii) measuring the quantity of the amplicons from the target sequences and amplicons from the 50 genetic standard sequences, and (iv) determining the relative amount of each of genetic standard sequences with respect to each other, whereby the plurality of genetic standard sequences is calibrated. In various embodiments, at least 10; 100, 500; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 55 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 genetic standard sequences are used. In various embodiments, the method involves using any of the primer libraries of the invention. In various embodiments, the method involves simultaneously ampli- 60 fying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different sequences. In some embodiments, the reference at chromosome 13, 18, 21, X, or Y. In some embodiments, the reference genome is diploid.

In one aspect, the invention provides a set of genetic standard sequences that have been calibrated according to any of the methods of the invention. In one aspect, the invention provides a set of genetic standard sequences that may be calibrated either before, during or after the method is performed.

In one aspect, the invention features methods of measuring the number of copies of a gene of interest having at least one allele that has a deletion. In some embodiments, the method includes (i) mixing genetic material derived from a sample for analysis with an amplification reagent specific for the gene of interest and not capable of significantly amplifying the deletion comprising allele of the gene of interest, a standard sequence corresponding to gene of interest, an amplification reagent specific for a reference sequence, and standard sequence corresponding to the reference sequence; (ii) amplifying the gene sequence of interest, the standard sequence corresponding to the gene of interest, the reference sequence, and the standard sequence correspond-15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 20 ing to the reference sequence to produce gene of interest amplicons, reference sequence amplicons, and standard sequence amplicons; and (iii) measuring the quantity of target amplicons and standard sequence amplicons produced. In some embodiments, the measuring of quantity is achieved by counting sequence reads. In some embodiments, the method further includes determining the estimated copy number of at least one chromosome in the sample from which the genetic library was derived, wherein the determination involves comparing the number of sequences of target amplicons with the number of sequences of a standard amplicons. In some embodiments, the standard sequences and the genetic library include universal priming sites capable of being primed by the same primer. In some embodiments, the relative amounts of each of the sequences have been calibrated with respect to reference genome. In various embodiments, at least 10; 100, 500; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 genetic standard sequences are used. In various embodiments, the method involves using any of the primer libraries of the invention. In various embodiments, the method involves simultaneously amplifying at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target regions. In some embodiments, the reference genome is diploid. In some embodiments, the sample for analysis is derived from blood.

In some embodiments of any of the aspects of the invention, preferentially enriching the DNA in the sample (e.g., the first sample) at the target loci (e.g., the plurality of polymorphic loci) includes obtaining a plurality of precircularized probes where each probe targets one of the loci (e.g., polymorphic loci), where the 3' and 5' end of the probes are preferably designed to hybridize to a region of DNA that is separated from the polymorphic site of the locus by a small number of bases, where the small number is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 to 25, 26 to 30, 31 to 60, or a combination thereof, hybridizing the pre-circularized probes to DNA from the sample (e.g., the first sample), filling the gap between the hybridized probe ends using DNA polymerase, circularizing the pre-circularized probe, and amplifying the circularized probe.

In some embodiments of any of the aspects of the invengenome has at least one aneuploidy, such as an aneuploidy 65 tion, the preferentially enriching the DNA at the target loci (e.g., the plurality of polymorphic loci) includes obtaining a plurality of ligation-mediated PCR probes where each PCR

## US 11,519,035 B2

3

probe targets one of the target loci (e.g., the polymorphic loci), and where the upstream and downstream PCR probes are designed to hybridize to a region of DNA on one strand of DNA that is preferably separated from the polymorphic site of the locus by a small number of bases, where the small number is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 to 25, 26 to 30, 31 to 60, or a combination thereof, hybridizing the ligation-mediated PCR probes to the DNA from the sample (e.g., the first sample), filling the gap between the ligation-mediated PCR probe ends using DNA 10 polymerase, ligating the ligation-mediated PCR probes, and amplifying the ligated ligation-mediated PCR probes.

In some embodiments of various aspects of the invention, preferentially enriching the DNA at the target loci (e.g., plurality of polymorphic loci) includes obtaining a plurality of hybrid capture probes that target the loci (e.g., the polymorphic loci), hybridizing the hybrid capture probes to the DNA in the sample (e.g., the first sample) and physically removing some or all of the unhybridized DNA from the sample (e.g., the first sample) of DNA.

In some embodiments of any of the aspects of the invention, the hybrid capture probes are designed to hybridize to a region that is flanking but not overlapping the polymorphic site. In some embodiments, the hybrid capture probes are designed to hybridize to a region that is flanking but not 25 overlapping the polymorphic site, and where the length of the flanking capture probe may be selected from the group consisting of less than about 120 bases, less than about 110 bases, less than about 100 bases, less than about 90 bases, less than about 80 bases, less than about 70 bases, less than 30 about 60 bases, less than about 50 bases, less than about 40 bases, less than about 30 bases, and less than about 25 bases. In some embodiments, the hybrid capture probes are designed to hybridize to a region that overlaps the polymorphic site, and where the plurality of hybrid capture probes 3 comprise at least two hybrid capture probes for each polymorphic loci, and where each hybrid capture probe is designed to be complementary to a different allele at that polymorphic locus.

In some embodiments of any of the aspects of the inven- 40 tion, preferentially enriching the DNA a plurality of polymorphic loci includes obtaining a plurality of inner forward primers where each primer targets one of the polymorphic loci, and where the 3' end of the inner forward primers are designed to hybridize to a region of DNA upstream from the 45 polymorphic site, and separated from the polymorphic site by a small number of bases, where the small number is selected from the group consisting of 1, 2, 3, 4, 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, or 31 to 60 base pairs, optionally obtaining a plurality of inner reverse primers 50 where each primer targets one of the polymorphic loci, and where the 3' end of the inner reverse primers are designed to hybridize to a region of DNA upstream from the polymorphic site, and separated from the polymorphic site by a small number of bases, where the small number is selected from 55 the group consisting of 1, 2, 3, 4, 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, or 31 to 60 base pairs, hybridizing the inner primers to the DNA, and amplifying the DNA using the polymerase chain reaction to form amplicons.

In some embodiments of any of the aspects of the invention, the method also includes obtaining a plurality of outer forward primers where each primer targets one of the target (e.g., polymorphic loci), and where the outer forward primers are designed to hybridize to the region of DNA upstream from the inner forward primer, optionally obtaining a plurality of outer reverse primers where each primer targets one of the target loci (e.g., polymorphic loci), and where the

32

outer reverse primers are designed to hybridize to the region of DNA immediately downstream from the inner reverse primer, hybridizing the first primers to the DNA, and amplifying the DNA using the polymerase chain reaction.

In some embodiments of any of the aspects of the invention, the method also includes obtaining a plurality of outer reverse primers where each primer targets one of the polymorphic loci, and where the outer reverse primers are designed to hybridize to the region of DNA immediately downstream from the inner reverse primer, optionally obtaining a plurality of outer forward primers where each primer targets one of the target loci (e.g., the polymorphic loci), and where the outer forward primers are designed to hybridize to the region of DNA upstream from the inner forward primer, hybridizing the first primers to the DNA, and amplifying the DNA using the polymerase chain reaction.

In some embodiments of any of the aspects of the invention, preparing the sample (e.g., the first sample) further includes appending universal adapters to the DNA in the sample (e.g., the first sample) and amplifying the DNA in the sample (e.g., the first sample) using the polymerase chain reaction. In some embodiments, at least a fraction of the amplicons that are amplified are less than 100 bp, less than 90 bp, less than 80 bp, less than 70 bp, less than 65 bp, less than 60 bp, less than 55 bp, less than 50 bp, or less than 45 bp, and where the fraction is 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 99%.

In some embodiments of any of the aspects of the invention, amplifying the DNA is done in one or a plurality of individual reaction volumes, and where each individual reaction volume contains more than 100 different forward and reverse primer pairs, more than 200 different forward and reverse primer pairs, more than 500 different forward and reverse primer pairs, more than 1,000 different forward and reverse primer pairs, more than 2,000 different forward and reverse primer pairs, more than 5,000 different forward and reverse primer pairs, more than 10,000 different forward and reverse primer pairs, more than 20,000 different forward and reverse primer pairs, more than 50,000 different forward and reverse primer pairs, or more than 100,000 different forward and reverse primer pairs. In various embodiments of any of the aspects of the invention, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primer pairs are used.

In some embodiments of any of the aspects of the invention, preparing the sample (e.g., the first sample) further comprises dividing the sample (e.g., the first sample) into a plurality of portions, and where the DNA in each portion is preferentially enriched at a subset of the target loci (e.g., plurality of polymorphic loci). In some embodiments, the inner primers are selected by identifying primer pairs likely to form undesired primer duplexes and removing from the plurality of primers at least one of the pair of primers identified as being likely to form undesired primer duplexes. In some embodiments, the inner primers contain a region that is designed to hybridize either upstream or downstream of the targeted locus (e.g., the polymorphic locus), and optionally contain a universal priming sequence designed to allow PCR amplification. In some embodiments, at least some of the primers additionally contain a random region that differs for each individual primer molecule. In some embodiments, at least some of the primers additionally contain a molecular barcode.

In some embodiments of any of the aspects of the invention, preferential enrichment results in average degree of

## US 11,519,035 B2

3

allelic bias between the prepared sample and the sample (e.g., the first sample) of a factor selected from the group consisting of no more than a factor of 2, no more than a factor of 1.5, no more than a factor of 1.2, no more than a factor of 1.1, no more than a factor of 1.05, no more than a factor of 1.02, no more than a factor of 1.002, no more than a factor of 1.005, no more than a factor of 1.002, no more than a factor of 1.001 and no more than a factor of 1.0001. In some embodiments, the plurality of polymorphic loci are SNPs. In some embodiments, measuring the DNA in the 10 prepared sample is done by sequencing.

In some embodiments, the nucleic acids in the sample are non-specifically amplified prior to amplification of the target loci (such as specific amplification of the target loci with a primer library of the invention). In some embodiments, the 15 non-specific amplification includes whole genome application (WGA), such as ligation-mediated PCR (LM-PCR), degenerate oligonucleotide primer PCR (DOP-PCR), or multiple displacement amplification (MDA). In some embodiments, the non-specific amplification includes universal PCR, such as adaptor-mediated universal PCR.

In some embodiments of any of the aspects of the invention, the target loci are present on the same nucleic acid of interest (e.g, the same chromosome or the same region of a chromosome). In some embodiments, at least some of the 25 target loci are present on different nucleic acids of interest (e.g, different chromosomes). In some embodiments, the nucleic acid sample includes fragmented or digested nucleic acids. In some embodiments, the nucleic acid sample includes DNA, such as genomic DNA, cDNA, cell-free 30 DNA (cfDNA), cell-free mitochondrial DNA (cf mDNA), cell-free DNA that originated from nuclear DNA (cf nDNA), cellular DNA, or mitochondrial DNA. In some embodiments, nucleic acid sample includes RNA, such as cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, 35 non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA. In some embodiments, the nucleic acid sample includes DNA from a single cell, 2 cells, 3 cells, 4 cells, 5 cells, 6 cells, 7 cells, 8 cells, 9 cell, 10 cells, or more than 10 cells. In some embodiments, the nucleic 40 acid sample is a blood or plasma sample that is substantially free of cells. In some embodiments, the nucleic acid sample includes or is derived from blood, plasma, saliva, semen, sperm, cell culture supernatant, mucus secretion, dental plaque, gastrointestinal tract tissue, stool, urine, hair, bone, 45 body fluids, tears, tissue, skin, fingernails, blastomeres, embryos, amniotic fluid, chorionic villus samples, bile, lymph, cervical mucus, or a forensic sample. In some embodiments, the target loci are segments of human nucleic acids. In some embodiments, the target loci are segments of 50 human nucleic acids found in the human genome. In some embodiments, the target loci comprise or consist of single nucleotide polymorphisms (SNPs). In some embodiments, the primers are DNA molecules.

In some embodiments of any of the aspects of the invention, the DNA in the sample (e.g., the first sample) originates from maternal plasma. In some embodiments, preparing the sample (e.g., the first sample) further comprises amplifying the DNA. In some embodiments, preparing the sample (e.g., the first sample) further comprises preferentially enriching 60 the DNA in the sample (e.g., the first sample) at the target loci (e.g., a plurality of polymorphic loci).

In various embodiments, the primer extension reaction or the polymerase chain reaction includes the addition of one or more nucleotides by a polymerase. In some embodiments, 65 greater than or equal to 5, 10, 20, 30, 40, 50, or 60 cycles of PCR are performed. In some embodiments, the amplification 34

of loci is performed using a polymerase (e.g., a DNA polymerase, RNA polymerase, or reverse transcriptase) with low 5'→3' exonuclease and/or low strand displacement activity. In some embodiments, a DNA polymerase is used produce DNA amplicons using DNA as a template. In some embodiments, a RNA polymerase is used produce RNA amplicons using DNA as a template. In some embodiments, a reverse transcriptase is used produce cDNA amplicons using RNA as a template.

In various embodiments, the primer extension reaction or the polymerase chain reaction does not include ligationmediated PCR. In various embodiments, the primer extension reaction or the polymerase chain reaction does not include the joining of two primers by a ligase. In various embodiments, the primers do not include Linked Inverted Probes (LIPs), which can also be called pre-circularized probes, pre-circularizing probes, circularizing probes, Padlock Probes, or Molecular Inversion Probes (MIPs). In some embodiments, the primers are not loopable primers. In some embodiments, the primers do not form a loop structure, for example, the primers do not comprise a 3' target specific portion, a stem (comprising a first loop forming region and a second loop forming region), and a loop portion. In various embodiments, the primer library, composition, kit, or method does not include an array (such as a microarray) or do no use an array (such as a microarray). In some embodiments, multiplex PCR and/or sequencing is performed without use of an array (such as a microarray). In some embodiments, the primer library, composition, kit, or method comprises a microarray.

In some embodiments, the primers or the target loci do not comprise an STR allele (for example, the target loci may be non-polymorphic loci or the loci may comprise a polymorphism other than an STR allele). In some embodiments, some or all of the target loci comprise an STR allele.

It is understood that all of the aspects and embodiments of the invention described herein include "comprising," "consisting," and "consisting essentially of" aspects and embodiments. It is understood that aspects and embodiments of the invention described herein include combinations of any two or more of the aspects or embodiments of the invention.

#### Definitions

Single Nucleotide Polymorphism (SNP) refers to a single nucleotide that may differ between the genomes of two members of the same species. The usage of the term should not imply any limit on the frequency with which each variant occurs.

Sequence refers to a DNA sequence or a genetic sequence. It may refer to the primary, physical structure of the DNA molecule or strand in an individual. It may refer to the sequence of nucleotides found in that DNA molecule, or the complementary strand to the DNA molecule. It may refer to the information contained in the DNA molecule as its representation in silico.

Locus refers to a particular region of interest on the DNA (or corresponding RNA) of an individual, which may refer to a SNP, the site of a possible insertion or deletion, or the site of some other relevant genetic variation. Disease-linked SNPs may also refer to disease-linked loci.

Polymorphic Allele, also "Polymorphic Locus," refers to an allele or locus where the genotype varies between individuals within a given species. Some examples of polymorphic alleles include single nucleotide polymorphisms, short tandem repeats, deletions, duplications, and inversions.

## US 11,519,035 B2

3:

Polymorphic Site refers to the specific nucleotides found in a polymorphic region that vary between individuals.

Allele refers to the alternative form or version of a gene that occupies a particular locus. Genetic Data also "Genotypic Data" refers to the data describing aspects of the genome of one or more individuals. It may refer to one or a set of loci, partial or entire sequences, partial or entire chromosomes, or the entire genome. It may refer to the identity of one or a plurality of nucleotides; it may refer to a set of sequential nucleotides, or nucleotides from different locations in the genome, or a combination thereof. Genotypic data is typically in silico, however, it is also possible to consider physical nucleotides in a sequence as chemically encoded genetic data. Genotypic Data may be said to be "on," "of," "at," "from" or "on" the individual(s). Genotypic Data may refer to output measurements from a genotyping platform where those measurements are made on genetic material.

Genetic Material also "Genetic Sample" refers to physical  $_{20}$  matter, such as tissue or blood, from one or more individuals comprising DNA or RNA

Noisy Genetic Data refers to genetic data with any of the following: allele dropouts, uncertain base pair measurements, incorrect base pair measurements, missing base pair 25 measurements, uncertain measurements of insertions or deletions, uncertain measurements of chromosome segment copy numbers, spurious signals, missing measurements, other errors, or combinations thereof.

Confidence refers to the statistical likelihood that the 30 called SNP, allele, set of alleles, ploidy call, or determined number of chromosome segment copies correctly represents the real genetic state of the individual.

Ploidy Calling, also "Chromosome Copy Number Calling," or "Copy Number Calling" (CNC), may refer to the act 35 of determining the quantity and/or chromosomal identity of one or more chromosomes present in a cell.

Aneuploidy refers to the state where the wrong number of chromosomes (e.g., the wrong number of full chromosomes or the wrong number of chromosome segments, such as the 40 presence of deletions or duplications of a chromosome segment) is present in a cell. In the case of a somatic human cell it may refer to the case where a cell does not contain 22 pairs of autosomal chromosomes and one pair of sex chromosomes. In the case of a human gamete, it may refer to the case where a cell does not contain one of each of the 23 chromosomes. In the case of a single chromosome type, it may refer to the case where more or less than two homologous but non-identical chromosome copies are present, or where there are two chromosome copies present that originate from the same parent. In some embodiments, the deletion of a chromosome segment is a microdeletion.

Ploidy State refers to the quantity and/or chromosomal identity of one or more chromosomes types in a cell.

Chromosome may refer to a single chromosome copy, 55 meaning a single molecule of DNA of which there are 46 in a normal somatic cell; an example is 'the maternally derived chromosome 18'. Chromosome may also refer to a chromosome type, of which there are 23 in a normal human somatic cell; an example is 'chromosome 18'.

Chromosomal Identity may refer to the referent chromosome number, i.e. the chromosome type. Normal humans have 22 types of numbered autosomal chromosome types, and two types of sex chromosomes. It may also refer to the parental origin of the chromosome. It may also refer to a 65 specific chromosome inherited from the parent. It may also refer to other identifying features of a chromosome.

36

The State of the Genetic Material or simply "Genetic State" may refer to the identity of a set of SNPs on the DNA, to the phased haplotypes of the genetic material, and to the sequence of the DNA, including insertions, deletions, repeats and mutations. It may also refer to the ploidy state of one or more chromosomes, chromosomal segments, or set of chromosomal segments.

Allelic Data refers to a set of genotypic data concerning a set of one or more alleles. It may refer to the phased, haplotypic data. It may refer to SNP identities, and it may refer to the sequence data of the DNA, including insertions, deletions, repeats and mutations. It may include the parental origin of each allele.

Allelic State refers to the actual state of the genes in a set of one or more alleles. It may refer to the actual state of the genes described by the allelic data.

Allelic Ratio or allele ratio, refers to the ratio between the amount of each allele at a locus that is present in a sample or in an individual. When the sample was measured by sequencing, the allelic ratio may refer to the ratio of sequence reads that map to each allele at the locus. When the sample was measured by an intensity based measurement method, the allele ratio may refer to the ratio of the amounts of each allele present at that locus as estimated by the measurement method.

Allele Count refers to the number of sequences that map to a particular locus, and if that locus is polymorphic, it refers to the number of sequences that map to each of the alleles. If each allele is counted in a binary fashion, then the allele count will be whole number. If the alleles are counted probabilistically, then the allele count can be a fractional number.

Allele Count Probability refers to the number of sequences that are likely to map to a particular locus or a set of alleles at a polymorphic locus, combined with the probability of the mapping. Note that allele counts are equivalent to allele count probabilities where the probability of the mapping for each counted sequence is binary (zero or one). In some embodiments, the allele count probabilities may be binary. In some embodiments, the allele count probabilities may be set to be equal to the DNA measurements.

Allelic Distribution, or 'allele count distribution' refers to the relative amount of each allele that is present for each locus in a set of loci. An allelic distribution can refer to an individual, to a sample, or to a set of measurements made on a sample. In the context of sequencing, the allelic distribution refers to the number or probable number of reads that map to a particular allele for each allele in a set of polymorphic loci. The allele measurements may be treated probabilistically, that is, the likelihood that a given allele is present for a give sequence read is a fraction between 0 and 1, or they may be treated in a binary fashion, that is, any given read is considered to be exactly zero or one copies of a particular allele.

Allelic Distribution Pattern refers to a set of different allele distributions for different parental contexts. Certain allelic distribution patterns may be indicative of certain ploidy states.

Allelic Bias refers to the degree to which the measured ratio of alleles at a heterozygous locus is different to the ratio that was present in the original sample, such as a sample of DNA. The degree of allelic bias at a particular locus is equal to the observed allelic ratio at that locus, as measured, divided by the ratio of alleles in the original DNA or RNA sample at that locus. Allelic bias may be defined to be greater than one, such that if the calculation of the degree of allelic bias returns a value, x, that is less than 1, then the degree of

## US 11,519,035 B2

3

allelic bias may be restated as 1/x. Allelic bias maybe due to amplification bias, purification bias, or some other phenomenon that affects different alleles differently.

Allelic imbalance for aneuploidy determinations, such as CNV determinations, refers to the difference between the frequencies of the alleles for a locus. It is an estimate of the difference in the copy of numbers of the homologs. Allelic imbalance can arise from the complete loss of an allele or from an increase in copy number of one allele relative to the other. Allelic imbalances can be detected by measuring the 10 proportion of one allele relative to the other in fluids or cells from individuals that are constitutionally heterozygous at a given locus. (Mei et al, Genome Res, 10:1126-37 (2000)). For dimorphic SNPs that have alleles arbitrarily designated 'A' and 'B', the allele ratio of the A allele is  $n_4/(n_4+n_B)$ , 15 where  $n_A$  and  $n_B$  are the number of sequencing reads for alleles A and B, respectively. Allelic imbalance is the difference between the allele ratios of A and B for loci that are heterozygous in the germline. This definition is analogous to that for SNVs, where the proportion of abnormal DNA is 20 typically measured using mutant allele frequency, or n<sub>m</sub>/  $(n_m+n_r)$ , where  $n_m$  and  $n_r$  are the number of sequencing reads for the mutant allele and the reference allele, respectively.

Accordingly, the proportion of abnormal DNA for a CNV can be measured by the average allelic imbalance (AAI), 25 defined as I(H1-H2)I/(H1+H2), where Hi is the average number of copies of homolog i in the sample and Hi/(H1+H2) is the fractional abundance, or homolog ratio, of homolog i. The maximum homolog ratio is the homolog ratio of the more abundant homolog.

Primer, also "PCR probe" refers to a single DNA molecule (a DNA oligomer) or a collection of DNA molecules (DNA oligomers) where the DNA molecules are identical, or nearly so, and where the primer contains a region that is designed to hybridize to a targeted locus (e.g., a targeted polymorphic locus or a nonpolymorphic locus), and may contain a priming sequence designed to allow PCR amplification. A primer may also contain a molecular barcode. A primer may contain a random region that differs for each individual molecule. The terms "test primer" and "candidate primer" are not meant to be limiting and may refer to any of the primers disclosed herein.

Library of primers refers to a population of two or more primers. In various embodiments, the library includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 45 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers. In various embodiments, the library includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 50 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primer pairs, wherein each pair of primers includes a forward test primer and a reverse test primer where each pair of test primers hybridize to a target locus. In some embodiments, the library of primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different individual primers that each hybridize to a different target locus, wherein the individual primers are not part of primer pairs. 60 In some embodiments, the library has both (i) primer pairs and (ii) individual primers (such as universal primers) that are not part of primer pairs.

Different primers refers to non-identical primers. Different pools refers to non-identical pools.

Different target loci refers to non-identical target loci. Different amplicons refers to non-identical amplicons.

38

Hybrid Capture Probe refers to any nucleic acid sequence, possibly modified, that is generated by various methods such as PCR or direct synthesis and intended to be complementary to one strand of a specific target DNA sequence in a sample. The exogenous hybrid capture probes may be added to a prepared sample and hybridized through a denature-reannealing process to form duplexes of exogenous-endogenous fragments. These duplexes may then be physically separated from the sample by various means.

Sequence Read refers to data representing a sequence of nucleotide bases that were measured using a clonal sequencing method. Clonal sequencing may produce sequence data representing single, or clones, or clusters of one original DNA molecule. A sequence read may also have associated quality score at each base position of the sequence indicating the probability that nucleotide has been called correctly.

Mapping a sequence read is the process of determining a sequence read's location of origin in the genome sequence of a particular organism. The location of origin of sequence reads is based on similarity of nucleotide sequence of the read and the genome sequence.

Matched Copy Error, also "Matching Chromosome Aneuploidy" (MCA), refers to a state of aneuploidy where one cell contains two identical or nearly identical chromosomes.

This type of aneuploidy may arise during the formation of the gametes in meiosis, and may be referred to as a meiotic non-disjunction error. This type of error may arise in mitosis. Matching trisomy may refer to the case where three copies of a given chromosome are present in an individual and two of the copies are identical.

Unmatched Copy Error, also "Unique Chromosome Aneuploidy" (UCA), refers to a state of aneuploidy where one cell contains two chromosomes that are from the same parent, and that may be homologous but not identical. This type of aneuploidy may arise during meiosis, and may be referred to as a meiotic error. Unmatching trisomy may refer to the case where three copies of a given chromosome are present in an individual and two of the copies are from the same parent, and are homologous, but are not identical. Note that unmatching trisomy may refer to the case where two homologous chromosomes from one parent are present, and where some segments of the chromosomes are identical while other segments are merely homologous.

Homologous Chromosomes refers to chromosome copies that contain the same set of genes that normally pair up during meiosis.

Identical Chromosomes refers to chromosome copies that contain the same set of genes, and for each gene they have the same set of alleles that are identical, or nearly identical.

Allele Drop Out (ADO) refers to the situation where at least one of the base pairs in a set of base pairs from homologous chromosomes at a given allele is not detected.

Locus Drop Out (LDO) refers to the situation where both base pairs in a set of base pairs from homologous chromosomes at a given allele are not detected.

Homozygous refers to having similar alleles as corresponding chromosomal loci.

Heterozygous refers to having dissimilar alleles as corresponding chromosomal loci.

Heterozygosity Rate refers to the rate of individuals in the population having heterozygous alleles at a given locus. The heterozygosity rate may also refer to the expected or measured ratio of alleles, at a given locus in an individual, or a sample of DNA.

Highly Informative Single Nucleotide Polymorphism (HISNP) refers to a SNP where the fetus has an allele that is not present in the mother's genotype.

## US 11,519,035 B2

39

Chromosomal Region refers to a segment of a chromosome, or a full chromosome.

Segment of a Chromosome refers to a section of a chromosome that can range in size from one base pair to the entire chromosome.

Chromosome refers to either a full chromosome, or a segment or section of a chromosome.

Copies refers to the number of copies of a chromosome segment. It may refer to identical copies, or to non-identical, homologous copies of a chromosome segment wherein the 10 different copies of the chromosome segment contain a substantially similar set of loci, and where one or more of the alleles are different. Note that in some cases of aneuploidy, such as the M2 copy error, it is possible to have some copies of the given chromosome segment that are identical as well 15 as some copies of the same chromosome segment that are not identical

Haplotype refers to a combination of alleles at multiple loci that are typically inherited together on the same chromosome. Haplotype may refer to as few as two loci or to an 20 entire chromosome depending on the number of recombination events that have occurred between a given set of loci. Haplotype can also refer to a set of single nucleotide polymorphisms (SNPs) on a single chromatid that are statistically associated.

Haplotypic Data, also "Phased Data" or "Ordered Genetic Data," refers to data from a single chromosome in a diploid or polyploid genome, i.e., either the segregated maternal or paternal copy of a chromosome in a diploid genome.

Phasing refers to the act of determining the haplotypic 30 genetic data of an individual given unordered, diploid (or polyploidy) genetic data. It may refer to the act of determining which of two genes at an allele, for a set of alleles found on one chromosome, are associated with each of the two homologous chromosomes in an individual.

Phased Data refers to genetic data where one or more haplotypes have been determined.

Hypothesis refers to a possible ploidy state at a given set of chromosomes, or a set of possible allelic states at a given set of loci. The set of possibilities may comprise one or more 40 elements.

Copy Number Hypothesis, also "Ploidy State Hypothesis," refers to a hypothesis concerning the number of copies of a chromosome in an individual. It may also refer to a hypothesis concerning the identity of each of the chromosomes, including the parent of origin of each chromosome, and which of the parent's two chromosomes are present in the individual. It may also refer to a hypothesis concerning which chromosomes, or chromosome segments, if any, from a related individual correspond genetically to a given chromosome from an individual.

Target Individual refers to the individual whose genetic state is being determined. In some embodiments, only a limited amount of DNA is available from the target individual. In some embodiments, the target individual is a fetus. 55 In some embodiments, there may be more than one target individual. In some embodiments, each fetus that originated from a pair of parents may be considered to be target individuals. In some embodiments, the genetic data that is being determined is one or a set of allele calls. In some 60 embodiments, the genetic data that is being determined is a ploidy call.

Related Individual refers to any individual who is genetically related to, and thus shares haplotype blocks with, the target individual. In one context, the related individual may 65 be a genetic parent of the target individual, or any genetic material derived from a parent, such as a sperm, a polar

40

body, an embryo, a fetus, or a child. It may also refer to a sibling, parent or a grandparent.

Sibling refers to any individual whose genetic parents are the same as the individual in question. In some embodiments, it may refer to a born child, an embryo, or a fetus, or one or more cells originating from a born child, an embryo, or a fetus. A sibling may also refer to a haploid individual that originates from one of the parents, such as a sperm, a polar body, or any other set of haplotypic genetic matter. An individual may be considered to be a sibling of itself.

Fetal refers to "of the fetus," or "of the region of the placenta that is genetically similar to the fetus". In a pregnant woman, some portion of the placenta is genetically similar to the fetus, and the free floating fetal DNA found in maternal blood may have originated from the portion of the placenta with a genotype that matches the fetus. Note that the genetic information in half of the chromosomes in a fetus is inherited from the mother of the fetus. In some embodiments, the DNA from these maternally inherited chromosomes that came from a fetal cell is considered to be "of fetal origin," not "of maternal origin."

DNA of Fetal Origin refers to DNA that was originally part of a cell whose genotype was essentially equivalent to that of the fetus.

DNA of Maternal Origin refers to DNA that was originally part of a cell whose genotype was essentially equivalent to that of the mother.

Child may refer to an embryo, a blastomere, or a fetus.

Note that in the presently disclosed embodiments, the concepts described apply equally well to individuals who are a born child, a fetus, an embryo or a set of cells therefrom. The use of the term child may simply be meant to connote that the individual referred to as the child is the genetic offspring of the parents.

Parent refers to the genetic mother or father of an individual. An individual typically has two parents, a mother and a father, though this may not necessarily be the case such as in genetic or chromosomal chimerism. A parent may be considered to be an individual.

Parental Context refers to the genetic state of a given SNP, on each of the two relevant chromosomes for one or both of the two parents of the target.

Develop As Desired, also "Develop Normally," refers to a viable embryo implanting in a uterus and resulting in a pregnancy, and/or to a pregnancy continuing and resulting in a live birth, and/or to a born child being free of chromosomal abnormalities, and/or to a born child being free of other undesired genetic conditions such as disease-linked genes. The term "develop as desired" is meant to encompass anything that may be desired by parents or healthcare facilitators. In some cases, "develop as desired" may refer to an unviable or viable embryo that is useful for medical research or other purposes.

Insertion into a Uterus refers to the process of transferring an embryo into the uterine cavity in the context of in vitro fertilization.

Maternal Plasma refers to the plasma portion of the blood from a female who is pregnant.

Clinical Decision refers to any decision to take or not take an action that has an outcome that affects the health or survival of an individual. In the context of prenatal diagnosis, a clinical decision may refer to a decision to abort or not abort a fetus. A clinical decision may also refer to a decision to conduct further testing, to take actions to mitigate an undesirable phenotype, or to take actions to prepare for the birth of a child with abnormalities.

## US 11,519,035 B2

4

Diagnostic Box refers to one or a combination of machines designed to perform one or a plurality of aspects of the methods disclosed herein. In an embodiment, the diagnostic box may be placed at a point of patient care. In an embodiment, the diagnostic box may perform targeted amplification followed by sequencing. In an embodiment the diagnostic box may function alone or with the help of a technician.

Informatics Based Method refers to a method that relies heavily on statistics to make sense of a large amount of data. In the context of prenatal diagnosis, it refers to a method designed to determine the ploidy state at one or more chromosomes or the allelic state at one or more alleles by statistically inferring the most likely state, rather than by directly physically measuring the state, given a large amount of genetic data, for example from a molecular array or sequencing. In an embodiment of the present disclosure, the informatics based technique may be one disclosed in this patent. In an embodiment of the present disclosure it may be 20 PARENTAL SUPPORTIM.

Primary Genetic Data refers to the analog intensity signals that are output by a genotyping platform. In the context of SNP arrays, primary genetic data refers to the intensity signals before any genotype calling has been done. In the context of sequencing, primary genetic data refers to the analog measurements, analogous to the chromatogram, that comes off the sequencer before the identity of any base pairs have been determined, and before the sequence has been mapped to the genome.

Secondary Genetic Data refers to processed genetic data that are output by a genotyping platform. In the context of a SNP array, the secondary genetic data refers to the allele calls made by software associated with the SNP array reader, wherein the software has made a call whether a given allele is present or not present in the sample. In the context of sequencing, the secondary genetic data refers to the base pair identities of the sequences have been determined, and possibly also where the sequences have been mapped to the genome.

Non-Invasive Prenatal Diagnosis (NPD), or also "Non-Invasive Prenatal Screening" (NPS), refers to a method of determining the genetic state of a fetus that is gestating in a mother using genetic material found in the mother's blood, where the genetic material is obtained by drawing the 45 mother's intravenous blood.

Preferential Enrichment of DNA that corresponds to a locus, or preferential enrichment of DNA at a locus, refers to any method that results in the percentage of molecules of DNA in a post-enrichment DNA mixture that correspond to 50 the locus being higher than the percentage of molecules of DNA in the pre-enrichment DNA mixture that correspond to the locus. The method may involve selective amplification of DNA molecules that correspond to a locus. The method may involve removing DNA molecules that do not corre- 55 spond to the locus. The method may involve a combination of methods. The degree of enrichment is defined as the percentage of molecules of DNA in the post-enrichment mixture that correspond to the locus divided by the percentage of molecules of DNA in the pre-enrichment mixture that 60 correspond to the locus. Preferential enrichment may be carried out at a plurality of loci. In some embodiments of the present disclosure, the degree of enrichment is greater than 20. In some embodiments of the present disclosure, the degree of enrichment is greater than 200. In some embodi- 65 ments of the present disclosure, the degree of enrichment is greater than 2,000. When preferential enrichment is carried

42

out at a plurality of loci, the degree of enrichment may refer to the average degree of enrichment of all of the loci in the set of loci

Amplification refers to a method that increases the number of copies of a molecule, such as a molecule of DNA.

Selective Amplification may refer to a method that increases the number of copies of a particular molecule of DNA, or molecules of DNA that correspond to a particular region of DNA. It may also refer to a method that increases the number of copies of a particular targeted molecule of DNA, or targeted region of DNA more than it increases non-targeted molecules or regions of DNA. Selective amplification may be a method of preferential enrichment.

Universal Priming Sequence refers to a DNA sequence that may be appended to a population of target DNA molecules, for example by ligation, PCR, or ligation mediated PCR. Once added to the population of target molecules, primers specific to the universal priming sequences can be used to amplify the target population using a single pair of amplification primers. Universal priming sequences are typically not related to the target sequences.

Universal Adapters, or 'ligation adaptors' or 'library tags' are DNA molecules containing a universal priming sequence that can be covalently linked to the 5-prime and 3-prime end of a population of target double stranded DNA molecules. The addition of the adapters provides universal priming sequences to the 5-prime and 3-prime end of the target population from which PCR amplification can take place, amplifying all molecules from the target population, using a single pair of amplification primers.

Targeting refers to a method used to selectively amplify or otherwise preferentially enrich those molecules of DNA that correspond to a set of loci, in a mixture of DNA.

Joint Distribution Model refers to a model that defines the probability of events defined in terms of multiple random variables, given a plurality of random variables defined on the same probability space, where the probabilities of the variable are linked. In some embodiments, the degenerate case where the probabilities of the variables are not linked may be used.

Percent identity in reference to nucleic acid sequences refers to the degree of sequence identity between nucleic acid sequences.

## BRIEF DESCRIPTION OF THE DRAWINGS

The presently disclosed embodiments will be further explained with reference to the attached drawings, wherein like structures are referred to by like numerals throughout the several views. The drawings shown are not necessarily to scale, with emphasis instead generally being placed upon illustrating the principles of the presently disclosed embodiments.

FIG. 1: Graphical representation of direct multiplexed mini-PCR method.

FIG. 2: Graphical representation of semi-nested mini-PCR method.

FIG. 3: Graphical representation of fully nested mini-PCR method.

FIG. 4: Graphical representation of hemi-nested mini-PCR method.

FIG. 5: Graphical representation of triply hemi-nested mini-PCR method.

FIG. 6: Graphical representation of one-sided nested mini-PCR method.

FIG. 7: Graphical representation of one-sided mini-PCR method.

Case: 24-1324 Document: 42-1 Page: 144 Filed: 03/18/2024

## US 11,519,035 B2

43

FIG. 8: Graphical representation of reverse semi-nested mini-PCR method.

FIG. 9: Some possible workflows for semi-nested methods.

FIG. 10: Graphical representation of looped ligation adap- 5 tors.

FIG. 11: Graphical representation of internally tagged primers

FIG. 12: An example of some primers with internal tags (SEQ ID NOs: 44,611-44,622, respectively in order of appearance).

FIG. 13: Graphical representation of a method using primers with a ligation adaptor binding region.

FIG. 14: Simulated ploidy call accuracies for counting  $_{15}$ method with two different analysis techniques.

FIG. 15: Ratio of two alleles for a plurality of SNPs in a cell line in Example 4.

FIG. 16: Ratio of two alleles for a plurality of SNPs in a cell line in Example 4 sorted by chromosome.

FIGS. 17A-17D: Ratio of two alleles for a plurality of SNPs in four pregnant women plasma samples, sorted by chromosome.

FIG. 18: Fraction of data that can be explained by binomial variance before and after data correction.

FIG. 19: Graph showing relative enrichment of fetal DNA in samples following a short library preparation protocol

FIG. 20: Depth of read graph comparing direct PCR and semi-nested methods.

FIG. 21: Comparison of depth of read for direct PCR of 30 three genomic samples.

FIG. 22: Comparison of depth of read for semi-nested mini-PCR of three samples.

FIG. 23: Comparison of depth of read for 1,200-plex and 9,600-plex reactions.

FIG. 24: Read count ratios for six cells at three chromo-

FIGS. 25A-25C: Allele ratios for two three-cell reactions (FIGS. 25B and 25C) and a third reaction run on 1 ng of genomic DNA at three chromosomes (FIG. 25A).

FIGS. 26A and 26B: Allele ratios for two single-cell reactions (FIGS. 26A and 26B) at three chromosomes.

FIG. 27: Comparison of two primer libraries showing the number of loci with a particular minor allele frequency that are targeted by each primer library.

FIG. 28A: Graph of the electrophoresis of PCR products. FIGS. 28B-28M are electropherograms of lanes 1-12, respectively, in FIG. 28A.

FIG. 29: Cartoon depiction of a method of the invention for the determination of a fetal aneuploidy (FIG. 29, step A). 50 Maternal and paternal genotype data (from blood or buccal swabs) and crossover frequency data from the HapMap database are utilized to generate (FIG. 29, step B) multiple independent hypotheses for each potential fetal ploidy state in silico (FIG. 29, step C). Each of these hypotheses is 55 expanded to include sub-hypotheses with take into consideration the different possible crossover points. The data model predicts what the sequencing data would look like (the expected allele distributions) given each hypothetical fetal genotype and at different fetal cfDNA fractions, and is 60 compared to the actual sequencing data; the likelihood for each hypothesis is determined using Bayesian statistics. In this hypothetical example, the hypotheses with the highest likelihoods (euploidy) are determined (FIG. 29, step D). The individual likelihoods from FIG. 29, step C are summed for 65 each copy number hypothesis family (monosomy, disomy, or triploidy). The hypothesis with the maximum likelihood

is called as the ploidy state, reveals the fetal fraction, and

represents the sample-specific calculated accuracy. FIGS. 30, 30D-H: Typical graphical representations of euploidy (FIG. 30), monosomy (FIG. 30D), and trisomy (FIGS. 30E-30H). For all plots, the x-axis represents the linear position of the individual polymorphic loci along each chromosome (as indicated below the plots), and the y-axis represents the number of A allele reads as a fraction of the total (A+B) allele reads. Maternal and fetal genotypes, as well as the position on the y-axis around which the bands are centered, are indicated to the right of the plots. If desired to facilitate visualization, the plots may be color-coded according to maternal genotype, such that red (filled circles as shown in FIGS. 30, 30D-H) indicates a maternal genotype of AA, blue (filled squares as shown in FIGS. 30, 30D-H) indicates a maternal genotype of BB, and green (open triangles as shown in FIGS. 30, 30D-H) indicates a maternal genotype of AB. If desired, maternal allele contributions may be indicated in color in the "Fetal Genotype" column. Allele contributions are indicated as maternal|fetal, such that alleles for which the mother is AA and the fetus is AB are indicated as AA|AB. FIG. 30, 0% FF plot: Generated plots when two chromosomes are present and the fetal cfDNA fraction is 0%. This plot is from a non-pregnant woman, and 25 thus represents the pattern when the genotype is entirely maternal. Allele clusters are thus centered around 1 (AA alleles), 0.5 (AB alleles), and 0 (BB alleles). FIG. 30, 12% FF plot: Generated plot when two chromosomes are present and the fetal fraction is 12%. The contribution of fetal alleles to the fraction of A allele reads shifts the position of some allele spots up or down along the y-axis, such that the bands are centered around 1 (AA|AA alleles), 0.94 (AA|AB alleles), 0.56 (ABIAA alleles), 0.50 (ABIAB alleles), 0.44 (ABIBB alleles), 0.06 (BBIAB alleles), and 0 (BBIBB alleles). FIG. 30, 26% FF Plot. Generated plot when two chromosomes are present and the fetal fraction is 26%. The pattern, including two filled circles and two filled square peripheral bands and a trio of central open triangle bands, is readily apparent. Bands are centered around 1 (AA|AA alleles), 0.87 (AA|AB alleles), 0.63 (AB|AA alleles), 0.50 (AB|AB alleles), 0.37 (AB|BB alleles), 0.13 (BB|AB alleles), and 0 (BB|BB alleles). FIG. 30D: Generated plot when one chromosome is present and the fetal fraction is 26%. The hallmark pattern of one external filled circles and one external filled square peripheral band as well as two central open triangle bands indicated maternally-inherited monosomy. Because the fetus only contributes a single allele (A or B) to the allele reads, the internal peripheral filled circles and filled square bands are not present, and the center trio of bands condenses into two bands. Bands that are centered around 1 (AAIA alleles), 0.57 (ABIA alleles), 0.43 (ABIB alleles), and 0 (BBIB alleles). FIG. 30E: Generated plot when three chromosomes are present and the fetal fraction is 27%. This pattern of two filled circles and filled square peripheral bands as well as two central open triangle bands indicates maternally-inherited meiotic trisomy. Bands are centered around 1 (AA|AAA alleles), 0.88 (AA|AAB alleles), 0.56 (ABIAAB alleles), 0.44 (ABIABB alleles), 0.12 (BB|ABB alleles), and 0 (BB|BBB alleles). FIG. 30F: Generated plot when three chromosomes are present and the fetal fraction is 14%. This pattern of three filled circles and three filled square peripheral bands, as well as two central open triangle bands, indicates paternally-inherited meiotic trisomy. Bands are centered around 1 (AA|AAA alleles), 0.93 (AA|AAB alleles), 0.87 (AA|ABB alleles), 0.60 (ABIAAA alleles), 0.53 (ABIAAB alleles), 0.47 (ABIABB alleles), 0.40 (ABIBBB alleles), 0.13 (BBIAAB alleles),

#### US 11,519,035 B2

44

0.07 (BB|ABB alleles), and 0 (BB|BBB alleles). FIG. 30G: Generated plot when three chromosomes are present and the fetal fraction is 35%. This pattern of two filled circles and two filled square peripheral bands and four central open triangle bands indicates maternally-inherited mitotic trisomy. Bands are centered around 1 (AA|AAA alleles), 0.85 (AAIAAB alleles), 0.72 (ABIAAA alleles), 0.57 (ABIAAB alleles), 0.43 (AB|ABB alleles), 0.28 (AB|BBB alleles), 0.15 (BB|ABB alleles), and 0 (BB|BBB alleles). FIG. 30H: Generated plot when three chromosomes are present and the 10 fetal fraction is 25%. This pattern of two filled circles and two filled square peripheral bands as well as four central open triangle bands indicates paternally-inherited mitotic trisomy. This pattern can be distinguished from that of maternally-inherited mitotic trisomy (as in FIG. 30G) by the 15 position of the internal peripheral bands. Specifically, bands are centered around 1 (AA|AAA alleles), 0.78 (AA|ABB alleles), 0.67 (AB|AAA alleles), 0.56 (AB|AAB alleles), 0.44 (AB|ABB alleles), 0.33 (AB|BBB alleles), 0.22 (BB|AAB alleles), and 0 (BB|BBB alleles).

FIGS. 31A-31G: Graphical representations of (FIG. 31A) euploid, (FIG. 31B) T13, (FIG. 31C) T18, (FIG. 31D) T21, (FIG. 31E) 45,X, (FIG. 31F) 47, XXY, and (FIG. 31G) 47, XYY, test samples as indicated. Each chromosome is indicated at the top of the plot, fetal and maternal genotypes are indicated to the right of the plots, the x-axis represents the linear position of the SNPs along each chromosome, and the y-axis indicates the number of A allele reads as a fraction of the total reads. Note the altered cluster positioning based on fetal fraction, as described herein. Each spot represents a single SNP locus. Fetal and maternal genotypes are indicated to the right of the plot, and chromosome identities are indicated at the top of the plots.

FIG. 32: The combined at-birth prevalence of sex chromosome aneuploidies is greater than that of autosomal 35 aneuploidies.

FIGS. 33A-33F: Illustrations of the calculation of an interaction score between primers in a primer library. FIG. 33A shows the first two bases (dinucleotide) of a primer that align to the other primer for calculation of  $\Delta G$ . Iterate over 40 the remainder of the primer that aligns with the other primer by sliding the bases being observing one base to the right. Continue until  $\Delta G$  has been calculated for all pairs of bases that align (FIG. 33B). Shift the alignment of the two primers (FIGS. 33C and 33D). Determine  $\Delta G$  for the new alignment 45 (FIGS. 33E and 33F).

FIG. 34: Table of the percentage of reads that map to target loci for genomic DNA samples and for samples of a single cell from a cell line for both mother and child samples

FIG. 35: Overlay of depth of read for a genomic and a single cell sample for different SNPs.

FIG. 36: Table of the percentage of reads that map to target loci for blastoceol fluid and for a single blastocyst cell.

FIG. 37. Graph of reference counts (counts of one allele, 55 such as the "A" allele) divided by total counts for that locus for a single blastocyst cell.

FIG. 38 is a graph showing the limit of detection for single nucleotide variants in a tumor biopsy using three different methods described in Example 23.

FIG. **39** is a graph showing the limit of detection for single nucleotide variants in a plasma sample using three different methods described in Example 23.

FIGS. **40**A and **40**B are graphs of the analysis of genomic DNA (FIG. **40**A) or DNA from a single cell (FIG. **40**B) 65 using a library of approximately 28,000 primers designed to detect CNVs. The presence of two central bands instead of

46

one central band indicates the presence of a CNV. The x-axis represents the linear position of the SNPs, and the y-axis indicates the fraction of A allele reads out of the total reads.

FIGS. **41**A and **41**B are graphs of the analysis of genomic DNA (FIG. **41**A) or DNA from a single cell (FIG. **41**B) using a library of approximately 3,000 primers designed to detect CNVs. The presence of two central bands instead of one central band indicates the presence of a CNV. The x-axis represents the linear position of the SNPs, and the y-axis indicates the fraction of A allele reads out of the total reads.

FIG. **42** is a graph illustrating the uniformity in depth of read (DOR) for these ~3,000 loci.

FIG. 43 is a table comparing error call metrics for genomic DNA and DNA from a single cell.

FIG. 44 is a graph of error rates for transition mutations and transversion mutations.

FIG. **45** is a table of data (such as percent mapped reads and error rate) from multiplex PCR with various buffers.

FIG. 46 is a graph illustrating the uniformity in DOR for multiplex PCR with buffers from FIG. 45.

FIG. 47 is a graph illustrating the normalized depth of read (DOR) for multiplex PCR with buffers from FIG. 45 with the DOR normalized to that of buffer 2×MM.

While the above-identified drawings set forth presently disclosed embodiments, other embodiments are also contemplated, as noted in the discussion. This disclosure presents illustrative embodiments by way of representation and not limitation. Numerous other modifications and embodiments can be devised by those skilled in the art which fall within the scope and spirit of the principles of the presently disclosed embodiments.

### DETAILED DESCRIPTION

The present invention is based in part on the surprising discovery that often only a relatively small number of primers in a library of primers are responsible for a substantial amount of the amplified primer dimers that form during multiplex PCR reactions. Methods have been developed to select the most undesirable primers for removal from a library of candidate primers. By reducing the amount of primer dimers to a negligible amount (~0.1% of the PCR products), these methods allow the resulting primer libraries to simultaneously amplify a large number of target loci in a single multiplex PCR reaction. Because the primers hybridize to the target loci and amplify them rather than hybridizing to other primers and forming amplified primer dimers, the number of different target loci that can be amplified is increased. It was also discovered that using lower primer concentrations and much longer annealing times than normal increases the likelihood that the primers hybridize to the target loci instead of hybridizing to each other and forming primer dimers (see, e.g., U.S. Ser. No. 13/683,604, filed Nov. 21, 2012, which is hereby incorporated by reference in its entirety). The methods can also be used to amplify a large number of target loci even if the primers have a relatively large range of melting temperatures (in contrast to other methods that require primers to have very similar melting temperatures to successfully amplify multiple loci simultaneously due to the need for the primers to be functional under the same reaction conditions). Additionally, it was discovered that the annealing temperature can optionally be higher than the melting temperatures of the primers (in contrast to other methods that use an annealing temperature below the melting temperatures of the primers). A higher

#### US 11,519,035 B2

4

annealing temperature improves the specificity of the PCR amplification and reduces or prevents amplification of non-target loci.

During the PCR amplification and sequencing of 19,488 target loci in a genomic sample, 99.4-99.7% of the sequencing reads mapped to the genome, of those, 99.99% of the mapped to target loci. For plasma samples with 10 million sequencing reads, typically at least 19,350 of the 19,488 target loci (99.3%) were amplified and sequenced. This primer library has even been used to amplify the nucleic 10 acids in a single cell (FIGS. **34-37**).

During the PCR amplification and sequencing of ~28,000 target loci in a genomic sample, 99% of the sequencing reads mapped to target loci. This primer library has also been used to amplify nucleic acids in a single cell.

Being able to simultaneously amplify such a large number of target loci at once greatly decreases the amount of time and the amount of DNA required to analyze thousands of target loci. For example, DNA from a single cell is sufficient to simultaneously analyze thousands of target loci, which is 20 important for applications in which the amount of DNA is low, such as genetic testing of a single cell from an embryo prior to in vitro fertilization or genetic testing of a forensic sample with little DNA. In addition, being able to analyze the target loci in one reaction volume (such as in one 25 chamber, well, or vessel) rather than splitting the sample into multiple different reactions reduces variability that can occur between reactions. In addition, methods have been developed to use reference standards to correct for amplification bias that may occur between different target loci. For 30 example, differences in amplification efficiency between target loci due to factors such as GC content may cause differing amounts of PCR products to be produced for target loci that are actually present in the same amount. The use of reference standards similar to the target loci allows the 3: detection of such amplification bias so that it can be corrected for during the quantitation of the target loci.

During sequencing of PCR products, artifacts such as primer dimers are detected and thus inhibit the detection of target amplicons. Because of this limitation, microarrays 4 with hybridization probes are often used for detection since microarrays are less sensitive to interference from primer dimers (for example, microarrays can be used as a target specific detection that uses probes to hybridize to target amplicons but does not have probes to hybridize to unde- 45 sired primer dimers). The high level of multiplexing with minimal non-target amplicons that has now been achieved allows PCR followed by sequencing to be used as an alternative to microarrays. For example, the present multiplex PCR methods can be used with a non-target specific 50 method of detection, such as sequencing that detects all amplified products (including both target amplicons and primer dimers, if any). The small amount of primer dimers that are produced allows detection of target amplicons by methods that detect all amplicons. Thus, in some embodi- 55 ments, the method includes multiplex PCR followed by sequencing without use of an array. In other embodiments, the method includes multiplex PCR followed by an array for detection of the amplified products.

The multiplex-PCR methods of the invention can be in a 60 variety of applications, such as genotyping, detection of chromosomal abnormalities (such as a fetal chromosome aneuploidy), gene mutation and polymorphism (such as single nucleotide polymorphisms, SNPs) analysis, gene deletion analysis, determination of paternity, analysis of 65 genetic differences among populations, forensic analysis, measuring predisposition to disease, quantitative analysis of

48

mRNA, and detection and identification of infectious agents (such as bacteria, parasite, and viruses). The multiplex PCR methods can also be used for non-invasive prenatal testing, such as paternity testing or the detection of fetal chromosome abnormalities.

Exemplary Primer Design Methods

Highly multiplexed PCR can often result in the production of a very high proportion of product DNA that results from unproductive side reactions such as primer dimer formation. In an embodiment, the particular primers that are most likely to cause unproductive side reactions may be removed from the primer library to give a primer library that will result in a greater proportion of amplified DNA that maps to the genome. The step of removing problematic primers, that is, those primers that are particularly likely to firm dimers has unexpectedly enabled extremely high PCR multiplexing levels for subsequent analysis by sequencing. In systems such as sequencing, where performance significantly degrades by primer dimers and/or other mischief products, greater than 10, greater than 50, and greater than 100 times higher multiplexing than other described multiplexing has been achieved. Note this is opposed to probe based detection methods, e.g. microarrays, TAQMAN, PCR etc. where an excess of primer dimers will not affect the outcome appreciably. Also note that the general belief in the art is that multiplexing PCR for sequencing is limited to about 100 assays in the same well. Fluidigm and Rain Dance offer platforms to perform 48 or 1000 s of PCR assays in parallel reactions for one sample.

There are a number of ways to choose primers for a library where the amount of non-mapping primer dimer or other primer mischief products are minimized. Empirical data indicate that a small number of 'bad' primers are responsible for a large amount of non-mapping primer dimer side reactions. Removing these 'bad' primers can increase the percent of sequence reads that map to targeted loci. One way to identify the 'bad' primers is to look at the sequencing data of DNA that was amplified by targeted amplification; those primer dimers that are seen with greatest frequency can be removed to give a primer library that is significantly less likely to result in side product DNA that does not map to the genome. There are also publicly available programs that can calculate the binding energy of various primer combinations, and removing those with the highest binding energy will also give a primer library that is significantly less likely to result in side product DNA that does not map to the genome

In some embodiments for selecting primers, an initial library of candidate primers is created by designing one or more primers or primer pairs to candidate target loci. A set of candidate target loci (such as SNPs) can selected based on publically available information about desired parameters for the target loci, such as frequency of the SNPs within a target population or the heterozygosity rate of the SNPs. In one embodiment, the PCR primers may be designed using program (the worldwide Primer3 primer3.sourceforge.net; libprimer3 release 2.2.3, which is hereby incorporated by reference in its entirety). If desired, the primers can be designed to anneal within a particular annealing temperature range, have a particular range of GC contents, have a particular size range, produce target amplicons in a particular size range, and/or have other parameter characteristics. Starting with multiple primers or primer pairs per candidate target locus increases the likelihood that a primer or prime pair will remain in the library for most or all of the target loci. In one embodiment, the selection criteria may require that at least one primer pair per target

#### US 11,519,035 B2

49

locus remains in the library. That way, most or all of the target loci will be amplified when using the final primer library. This is desirable for applications such as screening for deletions or duplications at a large number of locations in the genome or screening for a large number of sequences (such as polymorphisms or other mutations) associated with a disease or an increased risk for a disease. If a primer pair from the library would produces a target amplicon that overlaps with a target amplicon produced by another primer pair, one of the primer pairs may be removed from the library to prevent interference.

In some embodiments, a score such as an "undesirability score" (higher score representing least desirability) is calculated (such as calculation on a computer) for most or all of the possible combinations of two primers from a library 15 of candidate primers. In various embodiments, a score (such as an undesirability score) is calculated for at least 80, 90, 95, 98, 99, or 99.5% of the possible combinations of candidate primers in the library. Each score (such as an undesirability score) is based at least in part on the likeli- 20 herein hood of dimer formation between the two candidate primers. If desired, the score (such as the undesirability score) may also be based on one or more other parameters selected from the group consisting of heterozygosity rate of the target locus, disease prevalence associated with a sequence (e.g., a 25 polymorphism) at the target locus, disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, specificity of the candidate primer for the target locus, size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target ampli- 30 con, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target amplicon, size of the target amplicon, number of SNPs within the candidate primer, location of SNPs within the 35 candidate primer, distance from an end of the amplicon to the target bases within the amplicon, and the number of target loci in an amplicon. In some embodiments, the lower the number of SNPs with the candidate primer (such as 2, 1 or 0 SNPs) the better. In some embodiments, there are no 40 SNPs in the candidate primer. In some embodiments, SNPs (if any) are preferably not in the last 5 nucleotides in the 3' end of the candidate primer. In some embodiments, the target bases (the bases of interest in a target locus) are preferably near an end (the 3' or 5' end) of the amplicon; this 45 may improve the quality of sequencing data (since bases near the end of an amplicon are sequenced more accurately), and/or allow shorter sequencing reads to be performed. In some embodiments, a single amplicon includes 2 or more target loci (such as 2 or more nearby SNPs or variants). In 50 some embodiments, the specificity of the candidate primer for the target locus includes the likelihood that the candidate primer will mis-prime by binding and amplifying a locus other than the target locus it was designed to amplify. In some embodiments, one or more or all the candidate primers 55 that mis-prime are removed from the library. In some embodiments to increase the number of candidate primers to choose from, candidate primers that may mis-prime are not removed from the library. In some embodiments, the optimal melting temperature for selection of the candidate primers is 60 57° C. In some embodiments, the optimal size for selection of the candidate primers is a length of 24 nucleotides. In some embodiments, the optimal GC content for selection of the candidate primers is 50%. If multiple factors are considered, the score (such as the undesirability score) may be 65 calculated based on a weighted average of the various parameters. The parameters may be assigned different

50

weights based on their importance for the particular application that the primers will be used for. An exemplary score (such as an undesirability score) for a primer is shown below in which the parameters are weighted by various factors.

```
score=(1)(total number of targets-number of targets
covered)+(100)(number of SNPs in GC clamp)+
(10)(number of SNPs in primer binding site)+
(10)(number of similar primer pair designs)+
(0.1)(distance of target base from amplicon
end)+(0.1)(amplicon length)+(100)(interaction
score)
```

where interaction score=max  $(-1*\Delta G \text{ value})$  as described herein

Another exemplary score for a primer is shown below.

```
score=(100)(number of SNPs in GC clamp)+(10)
(number of SNPs in primer binding site)+(0.1)
(distance of target base from amplicon end)+
(0.1)(amplicon length)+(100)(interaction score)
```

where interaction score=max ( $-1*\Delta G$  value) as described herein

In some embodiments, the score for a primer pair is the worse score out of the scores for the two primers in the pair. An exemplary score (such as an undesirability score or the score in Example 20) for a pairs of designs (in which each design is one primer pair so that a pair of designs includes two primer pairs with a total of 4 primers) is shown below.

score=(10000000)(amplicon overlap)+(100)(distance between designs)+(1)(total number of targetsnumber of targets covered)+(100)(number of SNPs in GC clamp)+(10)(number of SNPs in primer binding site)+(10)(number of similar primer pair designs)+(0.1)(distance of target base from amplicon end)+(0.1)(amplicon length)+(100)(interaction score)

where interaction score=max  $(-1*\Delta G \text{ value})$  as described herein;

where amplicon overlap=overlap between the two amplicons formed by a pair of designs

In some embodiments, the score for a pair of designs is the worse score out of the scores for the four primers in the pair of designs.

In some embodiments, the primer with the highest score (such as the undesirability score) or any score representing least desirability is removed from the library. If the removed primer is a member of a primer pair that hybridizes to one target locus, then the other member of the primer pair may be removed from the library. The process of removing primers may be repeated as desired. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below a minimum threshold (such as any threshold for which the primers remaining in the library all have at least that level of desirability). In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number.

In various embodiments, after the score (such as the undesirability score) are calculated, the candidate primer that is part of the greatest number of combinations of two candidate primers with a score (such as an undesirability score) above a first minimum threshold (such as any threshold for which the primers remaining in the library all have at least that level of desirability) is removed from the library. This step ignores interactions equal to or below the first minimum threshold since these interactions are less significant. If the removed primer is a member of a primer pair that hybridizes to one target locus, then the other member of the

#### US 11,519,035 B2

primer pair may be removed from the library. The process of removing primers may be repeated as desired. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the first minimum threshold. If the number of candidate primers remaining in the library is higher than desired, the number of primers may be reduced by decreasing the first minimum threshold to a lower second minimum threshold (such as any threshold with a stricter cutoff than 10 the first minimum threshold so that some of the least desirable primers are removed from the library) and repeating the process of removing primers. If the number of candidate primers remaining in the library is lower than desired, the method can be continued by increasing the first 15 minimum threshold to a higher second minimum threshold (such as any threshold with a less strict cutoff than the first minimum threshold) and repeating the process of removing primers using the original candidate primer library, thereby allowing more of the candidate primers to remain in the 20 library. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the second minimum threshold, or until the number of candidate primers remain- 25 ing in the library is reduced to a desired number.

If desired, primer pairs that produce a target amplicon that overlaps with a target amplicon produced by another primer pair can be divided into separate amplification reactions. Multiple PCR amplification reactions may be desirable for 30 applications in which it is desirable to analyze all of the candidate target loci (instead of omitting candidate target loci from the analysis due to overlapping target amplicons).

In various embodiments of any of the aspects of the invention, the selection method selects candidate primers and divides them into different pools (e.g., 2, 3, 4, 5, 6, or more different pools). Each pool can be used to simultaneously amplify a large number of target loci (or a subset of target loci) in a single reaction volume. In some embodiments, a graph coloring algorithm is used to divide candidate primers into different pools. If desired, this method can be used to minimize the number of different pools needed to amplify most or all of the target loci.

In some embodiments, most or all of the target loci (such as at least 70, 80, 90, 92, 94, 96, 98, 99, or 100% of the target 45 loci) are amplified by at least 2, 3, 4, 5, 6, or more different pools. In some embodiments, most or all of the bases in the target loci (such as at least 70, 80, 90, 92, 94, 96, 98, 99, or 100% of the bases in the target loci) are amplified by at least 2,3, 4, 5, 6, or more different pools. In some embodiments, 50 most or all of the bases in the target loci (such as at least 70, 80, 90, 92, 94, 96, 98, 99, or 100% of the bases in the target loci) are amplified by at least 2, 3, 4, 5, 6, or more different primers or primer pairs in different pools. For example, a particular base in a target locus may be amplified by at least 55 2, 3, 4, 5, 6, or more different primers or primer pairs; wherein each different primer or primer pair is in a different pool. Using different primers or primer pairs to amplify each base allows multiple independent measurements of the base to be made, thereby increasing the accuracy of the method. 60 Dividing the different primers or primer pairs that amplify the same base into different pools prevents interference due to overlapping amplicons being formed by different primers or primer pairs

In one aspect, the invention features methods of selecting 65 test primers from a library of candidate primers to form 2 or more different primer pools. In various embodiments, the

selection involves (i) calculating on a computer a score (such as an undesirability score) for most or all of the possible combinations of two candidate primers from the library, wherein each score (such as an undesirability score) is based at least in part on the likelihood of dimer formation between the two candidate primers; (ii) removing the candidate primer with the highest or worst score (such as an undesirability score) from the library of candidate primers; and (iii) if the candidate primer removed in step (ii) is a member of a primer pair, then removing the other member of the primer pair from the library of candidate primers; and (iv) optionally repeating steps (ii) and (iii), thereby selecting a first pool. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below a minimum threshold for the first pool. In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number for the first pool. In some embodiments, after the first pool is selected those primers are removed from further consideration and steps of the method (such as steps (ii) and (iii)) are repeated with the remaining primers to select a second pool. If desired, this method may be repeated to select the desired number of primer pools. In some embodiments, the score is based at least in part on the current coverage of the bases in the target locus (such as the number of other primer pools that have a primer or primer pair that amplifies a particular base in the target locus).

In one aspect, the invention features methods of selecting test primers from a library of candidate primers to form 2 or more different primer pools. In various embodiments, the selection of test primers are selected from a library of candidate primers involves (i) calculating on a computer a score (such as an undesirability score) for most or all of the possible combinations of two candidate primers from the library, wherein each score (such as an undesirability score) is based at least in part on the likelihood of dimer formation between the two candidate primers; (ii) removing from the library of candidate primers the candidate primer that is part of the greatest number of combinations of two candidate primers with a score (such as an undesirability score) above a first minimum threshold; (iii) if the candidate primer removed in step (ii) is a member of a primer pair, then removing the other member of the primer pair from the library of candidate primers; and (iv) optionally repeating steps (ii) and (iii), thereby selecting a first pool. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the first minimum threshold for the first pool. In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is reduced to a desired number for the first pool. In various embodiments, the selection method involves further reducing the number of candidate primers remaining in the library by decreasing the first minimum threshold used in step (ii) to a lower second minimum threshold and optionally repeating steps (ii) and (iii). In some embodiments, the selection method involves increasing the first minimum threshold used in step (ii) to a higher second minimum threshold and optionally repeating steps (ii) and (iii). In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primer combinations remaining in the library are all equal to or below the second minimum threshold, or until the number of candidate primers remaining in the library is reduced to a

#### US 11,519,035 B2

5.

desired number for the first pool. In some embodiments, after the first pool is selected those primers are removed from further consideration and steps of the method (such as steps (ii) and (iii)) are repeated with the remaining primers to select a second pool. If desired, this method may be repeated to select the desired number of primer pools. In some embodiments, the score is based at least in part on the current coverage of the bases in the target locus (such as the number of other primer pools that have a primer or primer pair that amplifies a particular base in the target locus).

As discussed above, in some embodiments, a library is formed by starting with a library of candidate primers and removing primers until the primers remaining in the library have the desired characteristics for use as a final primer library.

In other embodiments, candidate primers are added to a library (such as a library starting with no primers) to form a library with the desired characteristics. In some embodiments, the candidate primer or primer pair with the most desirable score (such as the lowest undesirability score) is 20 added to a library (such as a library starting with no primers). The process of adding candidate primers may be repeated as desired. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primers that have not been added to the 25 library are all above a minimum threshold (such that all the candidate primers that have not been chosen for the library all have worse scores than the threshold). In some embodiments, the selection method is performed until the number of candidate primers that have been added to the library 30 reaches a desired number. The library of selected primers can then be used in any of the methods of the invention.

In some embodiments, most (such as at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5%) or all of the possible sets of two primer pairs (two primer pairs with a total of 4 3: primers) are considered for inclusion in a library. In some embodiments, the set of two different candidate primer pair with the most desirable score (such as the lowest undesirability score) is added to a first pool (such as a first pool starting with no primers). In some embodiments, the set of 4 two different candidate primer pairs with the next most desirable score is added to the first pool if it is connected to at most two sets of two different candidate primer pairs in the first pool. By "connected" for purposes of this step is meant that a single candidate primer pair in one set of two 45 different candidate primer pairs is the same as a single candidate primer pair in another set of two different candidate primer pairs. If the set of two different candidate primer pairs is connected to more than two sets, it may be added to a different pool than the first pool. The process of set of two 50 different candidate primer pair to pool(s) may be repeated as desired for the next set of two different candidate primer pairs with the next most desirable score. In some embodiments, the selection method is performed until the score (such as the undesirability score) for the candidate primers 55 that have not been added to the pool(s) are all above a minimum threshold (such that all the candidate primers that have not been chosen for the pool(s) all have worse scores than the threshold). In some embodiments, the selection method is performed until the number of candidate primers 60 that have been added to the pool(s) reaches a desired number. In some embodiments, the method involves storing designs in N number of doubly linked list data structures with the design pairs. N represents the current number of different primer pools. Initially, N=1, since there is only one 65 primer pool. In some embodiments, a second pool is only created if necessary to include the desired target loci or the

54

desired level of coverage of target loci. The library of selected primers can then be used in any of the methods of the invention.

In some embodiments, the minimum threshold, the first minimum threshold, or the second minimum threshold is an interaction score equal to or about 20, 18, 16, 14, 12, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 kcal/mol. In some embodiments, the interaction score is calculated as followed as described further herein:

Interaction score=max( $-1*\Delta G$  value); or

interaction score=max(-deltaG 2.0.8\*(-deltaG 1))

where

deltaG\_2=Gibbs energy (energy required to break the dimer) for a dimer that is extensible by PCR on both ends, i.e., the 3' end of each primer anneals to the other primer; and deltaG\_1=Gibbs energy for a dimer that is extensible by PCR on at least one end.

These selection methods minimize the number of candidate primers that have to be removed from the library to achieve the desired reduction in primer dimers. By removing a smaller number of candidate primers from the library, more (or all) of the target loci can be amplified using the resulting primer library.

Multiplexing large numbers of primers imposes considerable constraint on the assays that can be included. Assays that unintentionally interact result in spurious amplification products. The size constraints of miniPCR may result in further constraints. In an embodiment, it is possible to begin with a very large number of potential SNP targets (between about 500 to greater than 1 million) and attempt to design primers to amplify each SNP. Where primers can be designed it is possible to attempt to identify primer pairs likely to form spurious products by evaluating the likelihood of spurious primer duplex formation between all possible pairs of primers using published thermodynamic parameters for DNA duplex formation. Primer interactions may be ranked by a scoring function related to the interaction and primers with the worst interaction scores are eliminated until the number of primers desired is met. In cases where SNPs likely to be heterozygous are most useful, it is possible to also rank the list of assays and select the most heterozygous compatible assays. Experiments have validated that primers with high interaction scores are most likely to form primer dimers. At high multiplexing it is not possible to eliminate all spurious interactions, but it is essential to remove the primers or pairs of primers with the highest interaction scores in silico as they can dominate an entire reaction, greatly limiting amplification from intended targets. We have performed this procedure to create multiplex primer sets of up to and in some cases more than 10,000 primers. The improvement due to this procedure is substantial, enabling amplification of more than 80%, more than 90%, more than 95%, more than 98%, and even more than 99% on target products as determined by sequencing of all PCR products, as compared to 10% from a reaction in which the worst primers were not removed. When combined with a partial semi-nested approach as previously described, more than 90%, and even more than 95% of amplicons may map to the targeted sequences.

Note that there are other methods for determining which PCR probes are likely to form dimers. In an embodiment, analysis of a pool of DNA that has been amplified using a non-optimized set of primers may be sufficient to determine problematic primers. For example, analysis may be done using sequencing, and those dimers which are present in the

## US 11,519,035 B2

5:

greatest number are determined to be those most likely to form dimers, and may be removed.

This method has a number of potential application, for example to SNP genotyping, heterozygosity rate determination, copy number measurement, and other targeted sequencing applications. In an embodiment, the method of primer design may be used in combination with the miniper Method described elsewhere in this document. In some embodiments, the primer design method may be used as part of a massive multiplexed PCR method.

The use of tags on the primers may reduce amplification and sequencing of primer dimer products. In some embodiments, the primer contains an internal region that forms a loop structure with a tag. In particular embodiments, the primers include a 5' region that is specific for a target locus, 15 an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In some embodiments, the loop region may lie between two binding regions where the two binding regions are designed to bind to contiguous or neighboring regions of 20 template DNA. In various embodiments, the length of the 3' region is at least 7 nucleotides. In some embodiments, the length of the 3' region is between 7 and 20 nucleotides, such as between 7 to 15 nucleotides, or 7 to 10 nucleotides, inclusive. In various embodiments, the primers include a 5' 25 region that is not specific for a target locus (such as a tag or a universal primer binding site) followed by a region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. Tag-primers 30 can be used to shorten necessary target-specific sequences to below 20, below 15, below 12, and even below 10 base pairs. This can be serendipitous with standard primer design when the target sequence is fragmented within the primer binding site or, or it can be designed into the primer design. 3: Advantages of this method include: it increases the number of assays that can be designed for a certain maximal amplicon length, and it shortens the "non-informative" sequencing of primer sequence. It may also be used in combination with internal tagging (see elsewhere in this document).

In an embodiment, the relative amount of nonproductive products in the multiplexed targeted PCR amplification can be reduced by raising the annealing temperature. In cases where one is amplifying libraries with the same tag as the target specific primers, the annealing temperature can be 45 increased in comparison to the genomic DNA as the tags will contribute to the primer binding. In some embodiments we are using considerably lower primer concentrations than previously reported along with using longer annealing times than reported elsewhere. In some embodiments the anneal- 50 ing times may be longer than 3 minutes, longer than 5 minutes, longer than 8 minutes, longer than 10 minutes, longer than 15 minutes, longer than 20 minutes, longer than 30 minutes, longer than 60 minutes, longer than 120 minutes, longer than 240 minutes, longer than 480 minutes, and 55 even longer than 960 minutes. In an embodiment, longer annealing times are used than in previous reports, allowing lower primer concentrations. In various embodiments, longer than normal extension times are used, such as greater than 3, 5, 8, 10, or 15 minutes. In some embodiments, the 60 primer concentrations are as low as 50 nM, 20 nM, 10 nM, 5 nM, 1 nM, and lower than 1 uM. This surprisingly results in robust performance for highly multiplexed reactions, for example 1,000-plex reactions, 2,000-plex reactions, 5,000plex reactions, 10,000-plex reactions, 20,000-plex reactions, 65 50,000-plex reactions, and even 100,000-plex reactions. In an embodiment, the amplification uses one, two, three, four

-56

or five cycles run with long annealing times, followed by PCR cycles with more usual annealing times with tagged primers.

To select target locations, one may start with a pool of candidate primer pair designs and create a thermodynamic model of potentially adverse interactions between primer pairs, and then use the model to eliminate designs that are incompatible with other the designs in the pool.

In an embodiment, the invention features a method of decreasing the number of target loci (such as loci that may contain a polymorphism or mutation associated with a disease or disorder or an increased risk for a disease or disorder such as cancer) that need to be detected for a diagnosis and/or increasing the disease load that is detected (e.g., increasing the number of polymorphisms or mutations that are detected). In some embodiments, the method includes ranking (such as ranking from highest to lowest) loci by frequency or reoccurrence of a polymorphism or mutation (such as a single nucleotide variation, insertion, or deletion, or any of the other variations described herein) in each locus among subjects with the disease or disorder such as cancer. In some embodiments, PCR primers are designed to some or all of the loci. During selection of PCR primers for a library of primers, primers to loci that have a higher frequency or reoccurrence (higher ranking loci) are favored over those with a lower frequency or reoccurrence (lower ranking loci). In some embodiments, this parameter is included as one of the parameters in the calculation of the scores (such as the undesirability scores) described herein. If desired, primers (such as primers to high ranking loci) that are incompatible with other designs in the library can be included in a different PCR library/pool. In some embodiments, multiple libraries/pools (such as 2, 3, 4, 5 or more) are used in separate PCR reactions to enable amplification of all (or a majority) of the loci represented by all the libraries/ pools. In some embodiment, this method is continued until sufficient primers are included in one or more libraries/pools such that the primers, in aggregate, enable the desired disease load to be captured for the disease or disorder (e.g., such as by detection of at least 80, 85, 90, 95, or 99% of the

In some embodiments, the library of candidate primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers or different primer pairs. In some embodiments, only a relatively small number of candidate primers need to be removed from the library to achieve the desired reduction in primer dimers. In some embodiments, less than 30, 20, 15, 10, 5, or 2% of the candidate primers are removed from the library prior to use of the resulting library for multiplex PCR amplification of target loci. In some embodiments, a relatively large number of candidate primers are removed from the library to achieve the desired characteristics for the resulting library. In some embodiments, at least 20, 30, 40, 50, 60, 70, 80, or 90% of the candidate primers are removed from the library prior to use of the resulting library for multiplex PCR amplification of target loci. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers or different primer pairs remain in the library (after removal of some candidate primers from the library).

After the selection process, the primers remaining in the library may be used in any of the methods of the invention. Exemplary Methods for Determining Interaction Scores

#### US 11,519,035 B2

Exemplary methods of determining a  $\Delta G$  value (such as the change in Gibbs free energy for primer dimer formation) or an interaction score that indicates the likelihood of dimer formation between candidate primers are described below. In some embodiments, a thermodynamic Nearest-Neighbors 5 approach is used to calculate the likelihood of dimer formation between any two primers (see, e.g., Rahmann and Grafe (2004), "Mean and variance of the Gibbs free energy of oligonucleotides in the nearest neighbor model under varying conditions" Bioinformatics 20, 2928-2933; Allawi, 10 H. T. & SantaLucia, J., Jr. (1998), "Thermodynamics of Internal C-T Mismatches in DNA", Nucleic Acids Res. 26, 2694-2701; Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. (1999), "Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A-A, 15 C-C, G-G, and T-T Mismatches", Biochemistry 38, 3468-3477; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest-Neighbor Thermodynamics of Internal A-C Mismatches in DNA: Sequence Dependence and pH Effects", Biochemistry 37, 9435-9444.; Allawi, H. T. & SantaLucia, J., Jr. (1998), 20 interaction score can be calculated as follows. "Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA", Biochemistry 37, 2170-2179; and Allawi, H. T. & SantaLucia, J., Jr. (1997), "Thermodynamics and NMR of Internal G-T Mismatches in DNA" Biochemistry 36, 10581-10594; MultiPLX 2.1 (Kaplinski L, 25 Andreson R, Puurand T, Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005 Apr. 15; 21(8):1701-2, which are each hereby incorporated by reference in its entirety).

In some embodiments, the following steps are performed. 30

For each primer in a set of candidate primers, compare to every other candidate primer in the following way. Implement an ungapped thermodynamic alignment between the two primers, matching the 5' end of the first primer to the 3' 35 end of the second primer. Taking the first two bases (dinucleotide) that align to the other primer and vice versa, determine the  $\Delta H$  and  $\Delta S$  values for the dinucleotide in one primer hybridizing to the dinucleotide in the other primer (see the "AT" hybridizing to "GA" in FIG. 33A).  $\Delta$ H and  $\Delta$ S 40 values for various combinations of dinucleotides are known and can be determined, for example, using a thermodynamic look up table (such as the Unified NN model parameters according to Allawi and SantaLucia (1997) "Thermodynamics and NMR of internal G-T mismatches in DNA". Bio- 45 chemistry, 36: 10581-10594, which is hereby incorporated by reference in its entirety). Use the  $\Delta H$  and  $\Delta S$  values to calculate  $\Delta G$  for that interaction as follows or as described in any known equation for this.

 $\Delta G$ =(1000.0\* $\Delta H$ -(temperature\*( $\Delta S$ +0.368\*(num-Phosphates/2)\*log(saltConcentration))))/1000.0

In some embodiments, one or more of the following conditions are assumed for this calculation: temperature: of 60.0° C., primer concentration of 100 nM, or salt concen- 55 tration of 100 mM. In some embodiments, other conditions are assumed for this calculation, such as the conditions that will be used for multiplex PCR with the pool. Iterate over the remainder of the primer that aligns with the other primer by sliding the bases being observing one base to the right. 60 Continue until  $\Delta G$  has been calculated for all dinucleotides that align (FIG. 33B). The  $\Delta G$  for that alignment of the primer pair is the sum of the  $\Delta G$  values for the various dinucleotides.

Step 2

Shift the alignment of the two primers (FIGS. 33C and 33D)

58

Step 3

Repeat step 1 again for the new alignment (FIGS. 33E and 33F).

Step 4

After aligning all possible alignments between the two primers, determine the final  $\Delta G$  value and the interaction

In some embodiments, the  $\Delta G$  value for a combination of primers is the lowest  $\Delta G$  value (the lowest numerical value, which is indicative of the greatest likelihood of primer dimer formation) out of the  $\Delta G$  values for all possible alignments between the two primers. For example, if one alignment has a  $\Delta G$  value of -12 kcal/mol and another alignment has a  $\Delta G$ value of -2 kcal/mol then -12 kcal/mol (worse value) is used as the  $\Delta G$  value for that combination of primers

For algorithms such as the one in Example 16 in which it is easiest to rank primers based on assigning the worse combination of primers (those with the greatest likelihood of dimer formation) the highest interaction score, then the

Interaction score= $max(-1*\Delta G \text{ value})$ 

For example, if one alignment has a  $\Delta G$  value of -12kcal/mol and another alignment has a ΔG value of -2 kcal/mol, then 12 kcal/mol is used as the interaction score. In this case, the interaction score with the largest numerical positive number indicates the least desirable combination of primers due to the greatest likelihood of primer dimer formation.

In some embodiments, the interaction score is calculated as follows (this score weights the  $\Delta G$  values depending on the number of ends that a dimer can be extended by PCR).

 $interaction\_score=max(-deltaG\_2, 0.8*(-deltaG\_1))$ 

where

deltaG\_2=Gibbs energy (energy required to break the dimer) for a dimer that is extensible by PCR on both ends, i.e., the 3' end of each primer anneals to the other primer; and deltaG\_1=Gibbs energy for a dimer that is extensible by PCR on at least one end.

In some embodiments, deltaG\_2 is determined by performing steps 1-4 above for all the alignments in which a dimer is extensible by PCR on both ends. The deltaG\_2 value is the lowest  $\Delta G$  value (the lowest numerical value, which is indicative of the greatest likelihood of primer dimer formation) for all the alignments in which a dimer is extensible by PCR on both ends

In some embodiments, deltaG\_1 is determined by performing steps 1-4 above for all the alignments in which a dimer is extensible by PCR on at least one end (such as by PCR on one end or by PCR on both ends). The deltaG\_1 value is the lowest  $\Delta G$  value (the lowest numerical value, which is indicative of the greatest likelihood of primer dimer formation) for all the alignments in which a dimer is extensible by PCR on at least one end.

In some embodiments, possible loop structures or gaps in alignment between primers are also considered.

In some embodiments,  $\Delta G$  values from step 4 for each possible combination of two primers (each possible primer dimer) in a library are all equal to or greater than -20, -18, -16, -14, -12, -10, -9, -8, -7, -6, -5, -4, -3, -2, or -1kcal/mol. In some embodiments,  $\Delta G$  values from step 4 for at least 80, 85, 90, 92, 94, 96, 98, 99, or 100% of the primers in the library for possible combinations of that primer with 65 other primers in the library are all equal to or greater than -20, -18, -16, -14, -12, -10, -9, -8, -7, -6, -5, -4, -3, -2,or -1 kcal/mol. In some embodiments, possible combina-

#### US 11,519,035 B2

59

tions of two primers in a library that have positive  $\Delta G$  values are ignored since these values are indicative of no likelihood to for primer dimers. In some embodiments for the possible combination of two primers in a library that have negative  $\Delta G$  values, the  $\Delta G$  values are between -20 and 0 kcal/mol, 5 such as between -15 and 0 kcal/mol, -10 and 0 kcal/mol, -8 and 0 kcal/mol, -7 and 0 kcal/mol, -6 and 0 kcal/mol, -5.5 and 0 kcal/mol, -5 and 0 kcal/mol, -45 and 0 kcal/mol, -35 and 0 kcal/mol, -35 and 06 kcal/mol, -35 and 07 kcal/mol, -37 and 07 kcal/mol, -38 and 08 kcal/mol, -39 and 09 kcal/mol, -39 and -39 and

In some embodiments, the interaction scores from step 4 for each possible combination of two primers in a library are all equal to or less than 20, 18, 16, 14, 12, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 kcal/mol. In some embodiments, the interaction 15 scores from step 4 for at least 80, 85, 90, 92, 94, 96, 98, 99, or 100% of the primers in the library for possible combinations of that primer with other primers in the library are all equal to or less than 20, 18, 16, 14, 12, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 kcal/mol. In some embodiments, possible combi- 20 nation of two primers in a library that have negative interaction scores are ignored since these values are indicative of no likelihood to for primer dimers. In some embodiments for the possible combination of two primers in a library that have positive interaction scores, the interaction scores are 25 between 20 and 0 kcal/mol, such as between 15 and 0 kcal/mol, 10 and 0 kcal/mol, 8 and 0 kcal/mol, 7 and 0 keal/mol, 6 and 0 keal/mol, 5.5 and 0 keal/mol, 5 and 0 kcal/mol, 4.5 and 0 kcal/mol, 4 and 0 kcal/mol, 3.5 and 0 keal/mol, 3 and 0 keal/mol, 2.5 and 0 keal/mol, 2 and 0 30 kcal/mol, or 1.5 and 0 kcal/mol, inclusive.

In some embodiments, the score (such as the undesirability score) for candidate primers is based at least in part on the  $\Delta G$  value or the interaction score that indicates the likelihood of dimer formation between candidate primers as 35 calculated using any of these methods.

**Exemplary Primer Libraries** In one aspect, the invention features libraries of primers, such as primers selected from a library of candidate primers using any of the methods of the invention. In some embodi- 40 ments, the library includes primers that simultaneously hybridize (or are capable of simultaneously hybridizing) to or that simultaneously amplify (or are capable of simultaneously amplifying) at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 45 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci in one reaction volume. In various embodiments, the library includes primers that simultaneously amplify (or are capable of simultaneously amplifying) between 1,000 to 2,000; 2,000 to 5,000; 5,000 50 to 7,500; 7,500 to 10,000; 10,000 to 20,000; 20,000 to 25,000; 25,000 to 30,000; 30,000 to 40,000; 40,000 to 50,000; 50,000 to 75,000; or 75,000 to 100,000 different target loci in one reaction volume, inclusive. In various embodiments, the library includes primers that simultane- 55 ously amplify (or are capable of simultaneously amplifying) between 1,000 to 100,000 different target loci in one reaction volume, such as between 1,000 to 50,000; 1,000 to 30,000; 1,000 to 20,000; 1,000 to 10,000; 2,000 to 30,000; 2,000 to 20,000; 2,000 to 10,000; 5,000 to 30,000; 5,000 to 20,000; 60 or 5,000 to 10,000 different target loci, inclusive. In some embodiments, the library includes primers that simultaneously amplify (or are capable of simultaneously amplifying) the target loci in one reaction volume such that less than 60, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.5% of the 65 amplified products are primer dimers. The various embodiments, the amount of amplified products that are primer

60

dimers is between 0.5 to 60%, such as between 0.1 to 40%, 0.1 to 20%, 0.25 to 20%, 0.25 to 10%, 0.5 to 20%, 0.5 to 10%, 1 to 20%, or 1 to 10%, inclusive. In some embodiments, the primers simultaneously amplify (or are capable of simultaneously amplifying) the target loci in one reaction volume such that at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products are target amplicons. In various embodiments, the amount of amplified products that are target amplicons is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 98%, 90 to 99.5%, or 95 to 99.5%, inclusive. In some embodiments, the primers simultaneously amplify (or are capable of simultaneously amplifying) the target loci in one reaction volume such that at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified. In various embodiments, the amount target loci that are amplified is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 99%, 90 to 99.5%, 95 to 99.9%, or 98 to 99.99% inclusive. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, the library of primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 primer pairs, wherein each pair of primers includes a forward test primer and a reverse test primer where each pair of test primers hybridize to a target locus. In some embodiments, the library of primers includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 individual primers that each hybridize to a different target locus, wherein the individual primers are not part of

In some embodiments, the library includes primers (such as at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers or different primer pairs) that simultaneously amplify (or are capable of simultaneously amplifying) the target loci (such as at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci) in one reaction volume such that one or more of the following: (i) less than 60, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.5% of the amplified products are primer dimers, (ii) at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products are target amplicons, (iii) at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified, (iv) at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold, (v) at least 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the amplified products map to the human genome, or (vi) any combination thereof.

In some embodiments, the library includes at least 1,000 different primers or different primer pairs (such as at least 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers or different primer pairs) that simultaneously amplify (or are capable of simultaneously amplifying) at least 1,000 different target loci (such as at least 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci) in one reaction volume such that one or more of the following: (i) less than 60% of the amplified products are primer dimers and at least 40% of the

#### US 11,519,035 B2

6

amplified products are target amplicons, (ii) less than 40% of the amplified products are primer dimers and at least 60% of the amplified products are target amplicons, (iii) less than 20% of the amplified products are primer dimers and at least 80% of the amplified products are target amplicons, (iv) less 5 than 10% of the amplified products are primer dimers and at least 90% of the amplified products are target amplicons, or (v) less than 5% of the amplified products are primer dimers and at least 95% of the amplified products are target amplicons. In some embodiments for the amplification of nucleic 10 acids (such as DNA or RNA) from a single cell (such as a fetal or embryonic cell), the library includes at least 1,000 different primers or different primer pairs (such as at least 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 15 100,000 different primers or different primer pairs) that simultaneously amplify (or are capable of simultaneously amplifying) at least 1,000 different target loci (such as at least 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 20 100,000 different target loci) in one reaction volume such that one or more of the following: (i) less than 60% of the amplified products are primer dimers and at least 10% of the amplified products are target amplicons, (ii) less than 40% of the amplified products are primer dimers and at least 10% of 25 the amplified products are target amplicons, (iii) less than 20% of the amplified products are primer dimers and at least 10% of the amplified products are target amplicons, (iv) less than 10% of the amplified products are primer dimers and at least 10% of the amplified products are target amplicons, (v) 30 less than 5% of the amplified products are primer dimers and at least 15% of the amplified products are target amplicons; (vi) less than 60% of the amplified products are primer dimers and at least 20% of the amplified products are target amplicons, (vii) less than 40% of the amplified products are 3: primer dimers and at least 20% of the amplified products are target amplicons, (viii) less than 20% of the amplified products are primer dimers and at least 20% of the amplified products are target amplicons, (ix) less than 10% of the amplified products are primer dimers and at least 20% of the 40 amplified products are target amplicons, (x) less than 5% of the amplified products are primer dimers and at least 20% of the amplified products are target amplicons; (xi) less than 60% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (xii) less 45 than 40% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (xiii) less than 20% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (xiv) less than 10% of the amplified products are 50 primer dimers and at least 40% of the amplified products are target amplicons, (xv) less than 5% of the amplified products are primer dimers and at least 45% of the amplified products are target amplicons; (xvi) less than 40% of the amplified products are primer dimers and at least 60% of the amplified 55 products are target amplicons, (xvii) less than 20% of the amplified products are primer dimers and at least 80% of the amplified products are target amplicons, (xviii) less than 10% of the amplified products are primer dimers and at least 90% of the amplified products are target amplicons, or 60 (xviiii) less than 5% of the amplified products are primer dimers and at least 95% of the amplified products are target amplicons. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 65 400-fold. In some embodiments, one or more of these embodiments (e.g., percent of primer dimers, target ampli62

cons, or amplified target loci) is achieved after greater than or equal to 5, 10, 20, 30, 40, 50, or 60 cycles of PCR are performed. In some embodiments for a library that amplifies human target loci, at least 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the amplified products map to the human genome.

In various embodiments, the concentration of each primer is less than 100, 75, 50, 25, 20, 10, 5, 2, 1, 0.5, 0.1, or 0.05 nM, or less than 500, 100, 10, or 1 uM. In various embodiments, the concentration of each primer is between 1 uM to 100 nM, such as between 1 uM to 1 nM, 1 to 75 nM, 1 to 20 nM, 1 to 10 nM, 2 to 50 nM or 5 to 50 nM, inclusive. In some embodiments, the concentration of one or more universal primers is between 0.2 to 3 μM, such as between 0.5 and 2.5  $\mu$ M, 0.5 to 1  $\mu$ M, or 1 to 2.5  $\mu$ M per primer, inclusive, and the concentration of each primer except the universal primer(s) is between 1 uM to 100 nM, such as between 1 uM to 1 nM, 1 to 75 nM, 1 to 20 nM, 1 to 10 nM, 2 to 50 nM or 5 to 50 nM, inclusive. In various embodiments, the GC content of the primers is between 30 to 80%, such as between 20 to 70%, 40 to 70%, or 50 to 60%, inclusive. In some embodiments, the range of GC content of the primers is less than 30, 20, 10, or 5%. In some embodiments, the range of GC content of the primers is between 5 to 30%, such as 5 to 20% or 5 to 10%, inclusive. In some embodiments, there is a high GC content in the 3' end of the primers. In some embodiments, there are at least 2 (such as 3, 4, or 5) guanines or cytosines in the last 5 bases at the 3' end of the primers. In some embodiments, there are at least 1 (such as 2 or 3) guanines or cytosines in the last 3 bases at the 3' end of the primers. In some embodiments, a maximum of 2 (such as 2, 1, or 0) bases in the last 5 bases at the 3' end of the primers are guanines or cytosines. In some embodiments, a maximum of 1 (such as 1 or 0) base in the last 5 bases at the 3' end of the primers is a guanine or cytosine. In some embodiments, the maximum length of a homopolymer (the same base in a row) in the primers is 12, 10, 8, 6, 5, 4, 3, or 2 consecutive nucleotides. In some embodiments, the melting temperature  $(T_m)$  of the test primers is between 40 to 80° C., such as 50 to 70° C., 55 to 65° C., 54 to 60.5° C., or 57 to 60.5° C., inclusive. In some embodiments, the  $T_m$  is calculated using the Primer3 pro $gram \ (libprimer 3 \ release \ 2.2.3) \ using \ the \ built-in \ Santa Lucia$ parameters World Wide primer3.sourceforge.net). In some embodiments, the range of melting temperature of the primers is less than 15, 10, 5, 3, or 1° C. In some embodiments, the range of melting temperature of the primers is between 1 to 15° C., such as between 1 to 10° C., 1 to 5° C., or 1 to 3° C., inclusive. In some embodiments, the range of melting temperatures of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the primers is between 1 to 15° C., such as between 1 to 10° C., 1 to 5° C., 1 to 3° C., 2 to 5° C., 3 to 10° C., or 3 to 5° C., inclusive. In some embodiments, the length of the primers is between 15 to 100 nucleotides, such as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, 20 to 65 nucleotides, inclusive. In some embodiments, the range of the length of the primers is less than 50, 40, 30, 20, 10, or 5 nucleotides. In some embodiments, the range of the length of the primers is between 5 to 50 nucleotides, such as 5 to 40 nucleotides, 5to 20 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodiCase: 24-1324 Document: 42-1 Page: 154 Filed: 03/18/2024

#### US 11,519,035 B2

ments, the length of the target amplicons is between 30 and 400 nucleotides, such as between 30 and 200 nucleotides, or 100 and 400 nucleotides, inclusive. In some embodiments, the length of the target amplicons is at least 100; 200; 300; 400; 500; 600; 700; 800; 900; 1,000; 1,200; 1,500; 2,000; or 5 3,000 nucleotides. In some embodiments, the length of the target amplicons is between 100 and 1,500 nucleotides, such as between 100 to 1,000; 100 to 500, 500 to 750, or 750 to 1,000 nucleotides, inclusive. Longer amplicons may be desirable, e.g., for applications in which is it desirable to 10 screen for multiple potential mutations in one amplicon, such as carrier screening. In some embodiments, one round of PCR is performed to produce relatively long amplicons (such as at least 250 or 500 nucleotides in length) and then a second round of PCR is performed to produce shorter 15 amplicons (to amplify regions within the amplicons amplified in the first round of PCR, such as regions of less than 200 or 100 nucleotides in length). In some embodiments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the range of 20 the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 or all of the target amplicons have a length that falls within the range of the average length of the amplicons±5% of the average length, average length±20%, average length±20%, or average length±30%, or average length±50%.

In some embodiments, library includes at least at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers that each includes a target specific sequence, such as a 3: sequence that binds a target locus but does not substantially bind to other nucleic acids (such as non-target loci) in a sample, e.g., a biological sample, which naturally includes other nucleic acids. In some embodiments, each primer binds and amplifies a target locus by at least 2, 4, 6, 8, 10, 40 20, 50-fold or more than one or more (or all) other nucleic acids (such as non-target loci) in a sample. In some embodiments, the library includes at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 45 75,000; or 100,000 different target specific primers (e.g., primers that are specific for a target locus). In some embodiments, part or all of the polynucleotide sequence is a non-random sequence for at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 50 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers in the library. In some embodiments, library also includes a universal primer, a random primer, and a primer with a non-naturally occurring polynucleotide sequence, or a primer with a polynucleotide 55 sequence not naturally found in a human in some embodiments, the universal or random primer has a non-naturally occurring polynucleotide sequence or a polynucleotide sequence not naturally found in a human.

In some embodiments, the composition includes at least 60 one primer (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers) with a polynucleotide sequence of a human nucleic acid and at least one primer 65 (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000;

27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers) with a polynucleotide sequence that is not found in a human (such as a universal primer, a primer that comprises a region or consists entirely of random nucleotides, or a primer with a region such as a tag or barcode of one or more nucleotides that are not found in a human or are not found in nature as part of the polynucleotide sequence of the primer). In some embodiments, at least one primer (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers) includes a region of one or more nucleotides that is not naturally part of the primer sequence (such as a region added to the 5' end of the target specific sequence in the primer or an internal region added between the 5' and 3' ends of the primer). In some embodiments, the primer is free of the nucleic acids (such as genes) which, in the naturally-occurring genome of the organism from which the nucleic acid of the invention is derived, flank the gene. In some embodiments, the primer has been separated from one or more components that naturally accompany the corresponding sequence in nature (such as in the genome of a human). Typically, each primer is at least 90, 95, 98, 99, 99.9, or 100%, by weight, free from the mol-7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 25 ecules (such as proteins, nucleic acids, and naturally-occurring organic molecules) that naturally accompany the corresponding sequence in nature (such as in the genome of a human). Purity can be assayed by any appropriate method, e.g., by electrophoresis or HPLC analysis.

In some embodiments, the primers in the library are not immobilized (such as not immobilized to a solid support) or not part of a microarray. In some embodiments, the primers are dissolved in solution (such as dissolved in the liquid phase). In some embodiments, the library comprises a microarray. In some embodiments, the amplified products are detected using an array, such as an array with probes to one or more chromosomes of interest (e.g., chromosome 13, 18, 21, X, Y, or any combination thereof).

In some embodiments, at least one of the primers (such as at least 20, 40, 80, 90, 95, 98, 99, 99.5, or 100% of the primers) in a library are nucleic acid analogs that have a lower likelihood of primer dimerization compared to the naturally-occurring nucleic acids (see, e.g., U.S. Pat. Nos. 7,414,118 and 6,001,611; which are each hereby incorporated by reference in its entirety). Exemplary nucleic acid analogs have a modified pyrimidine nucleobase, or a purine or pyrimidine base that contains an exocyclic amine.

In some embodiments, the primer library includes a small number of primers (such as less than 5, 2, 1, or 0.5% of the primers in the library) that do not have one or more of the properties described herein. In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have one or more of the following properties: (i)  $\Delta G$  values for possible combinations of that primer with other primers in the library are all equal to or greater than -20, -18, -16, -14, -12, -10, -9, -8, -7, -6, -5, -4, -3, -2,or –1 kcal/mol; (ii) ΔG values for the possible combination of that primer with other primers in the library that have negative  $\Delta G$  are between -20 and 0 kcal/mol, such as between -15 and 0 kcal/mol, -10 and 0 kcal/mol, -8 and 0 kcal/mol, -7 and 0 kcal/mol, -6 and 0 kcal/mol, -5.5 and 0 kcal/mol, -5 and 0 kcal/mol, -4.5 and 0 kcal/mol, -4 and 0 kcal/mol. -3.5 and 0 kcal/mol. -3 and 0 kcal/mol. -2.5 and 0 kcal/mol, -2 and 0 kcal/mol, or -1.5 and 0 kcal/mol, inclusive; (iii) the GC content is between 30 to 80%, such as between 20 to 70%, 40 to 70%, or 50 to 60%, inclusive; (iv) the range of GC content is less than 30, 20, 10, or 5% or the

#### US 11,519,035 B2

65

range of GC content of the primers is between 5 to 30%, such as 5 to 20%, or 5 to 10%, inclusive; (v) a maximum of 2 (such as 2, 1, or 0) bases in the last 5 bases at the 3' end of the primers are guanines or cytosines; (vi) the melting temperature  $(T_m)$  of the primers is between 40 to 80° C., 5 such as 50 to 70° C., 55 to 65° C., 54 to 60.5° C., or 57 to 60.5° C., inclusive; (vii) the range of melting temperature of the primers is less than 15, 10, 5, 3, or 1° C.; (viii) the range of melting temperature of the primers is between 1 to 15° C. such as between 1 to 10° C., 1 to 5° C., 1 to 3° C., 2 to 5° C., 3 to 10° C., or 3 to 5° C., inclusive; (ix) the length of the primers is between 15 to 100 nucleotides, such as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, or 20 to 65 nucleotides, inclusive; (x) the range of the length of the primers is less than 15 50, 40, 30, 20, 10, or 5 nucleotides; (xi) the range of the length of the primers is between 5 to 50 nucleotides, such as 5 to 40 nucleotides, 5 to 20 nucleotides, or 5 to 10 nucleotides, (xii) the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, 20 or 60 to 75 nucleotides; (xiii) the length of the target amplicons is between 30 and 400 nucleotides, such as between 30 and 200 nucleotides, or 100 and 400 nucleotides; (xiv) the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides; (xv) the range of the 25 length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides; (xvi) the maximum length of a homopolymer in the primers is 12, 10, 8, 6, 5, 4, 3, or 2 consecutive nucleotides; (xvii) the concentration of each primer is less 30 than 100, 75, 50, 25, 20, 10, 5, 2, 1, 0.5, 0.1, or 0.05 nM, or less than 500, 100, 10, or 1 uM; (xviii) the concentration of each primer is between 1 uM to 100 nM, such as between 1 uM to 1 nM, 1 to 75 nM, 1 to 20 nM, 1 to 10 nM, 2 to 50 nM, or 5 to 50 nM, inclusive; (xix) at least 80, 90, 92, 94, 35 96, 98, 99, or 100% of the molecules of that primer are extended to form amplified products; (xx) SNPs (if any) are not in the last 5 nucleotides in the 3' end of the candidate primer; (xxi) the target bases (the bases of interest in a target locus) are near an end (the 3' or 5' end) of the amplicon; 40 (xxii) the region of hybridization is separated from the polymorphic site by a small number of bases, where the small number is selected from the group consisting of 1, 2, 3, 4, 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, and 31 to 60; (xxiii) the length of the annealing step is greater 45 than 1, 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes, (xxiv) the length of the annealing step (per PCR cycle) is between 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive, (xxv) the length of the annealing step is greater than 5 minutes (such 50 greater than 10, or 15 minutes), and the concentration of each primer is less than 20 nM, (xxvi) the length of the annealing step is greater than 5 minutes (such greater than 10, or 15 minutes), and the concentration of each primer is between 1 to 20 nM, or 1 to 10 nM, inclusive; (xxvii) the 55 length of the annealing step is greater than 20 minutes (such as greater than 30, 45, 60, or 90 minutes), and the concentration of each primer is less than 1 nM; (xxviii) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or  $15^{\circ}$  C. greater than the melting temperature 60(such as the empirically measured or calculated T<sub>m</sub>) of the primers; (xxix) the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the melting temperature (such as the empirically mea- 65 sured or calculated  $T_m$ ) of the primers; (xxx) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,

66

or 15° C. greater than the highest melting temperature of the primers; (xxxi) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the average melting temperature of the primers, (xxxii) the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to  $15^{\circ}$  C., inclusive) greater than the average melting temperature of the primers; and (xxviii) any combination thereof. In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have one or more of the following properties: (i)  $\Delta G$ values for possible combinations of that primer with other primers in the library are all equal to or greater than -20, -18, -16, -14, -12, -10, -9, -8, -7, -6, -5, -4, -3, -2, or -1 kcal/mol; (ii) ΔG values for the possible combination of that primer with other primers in the library that have negative  $\Delta G$  are between -20 and 0 kcal/mol, such as between -15 and 0 kcal/mol, -10 and 0 kcal/mol, -8 and 0 kcal/mol, -7 and 0 kcal/mol, -6 and 0 kcal/mol, -5.5 and 0 kcal/mol, -5 and 0 kcal/mol, -4.5 and 0 kcal/mol, -4 and 0 kcal/mol, -3.5 and 0 kcal/mol, -3 and 0 kcal/mol, -2.5 and 0 kcal/mol, -2 and 0 kcal/mol, or -1.5 and 0 kcal/mol, inclusive; (iii) the melting temperature  $(T_m)$  of the primers is between 40 to 80° C., such as 50 to 70° C., 55 to 65° C. 54 to 60.5° C., or 57 to 60.5° C., inclusive; (iv) the range of melting temperature of the primers is less than 15, 10, 5, 3, or 1° C.; (v) the range of melting temperature of the primers is between 1 to 15° C., such as between 1 to 10° C., 1 to 5° C., 1 to 3° C., 2 to 5° C., 3 to 10° C., or 3 to 5° C., inclusive; (vi) the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides; (vii) the length of the target amplicons is between 30 and 400 nucleotides, such as between 30 and 200 nucleotides, or 100 and 400 nucleotides; (viii) the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides; (ix) the range of the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides; (x) the concentration of each primer is less than 100, 75, 50, 25, 20, 10, 5, 2, 1, 0.5, 0.1, or 0.05 nM, or less than 500, 100, 10, or 1 uM; (xi) the concentration of each primer is between 1 uM to 100 nM, such as between 1 uM to 1 nM, 1 to 75 nM. 1 to  $20\,\text{nM},\,1$  to  $10\,\text{nM},\,2$  to  $50\,\text{nM},\,\text{or}\,5$  to  $50\,\text{nM},\,\text{inclusive};$ (xii) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers; (xiii) the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15 $^{\circ}$  C., inclusive) greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers; (xiv) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the highest melting temperature of the primers; (xv) the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the average melting temperature of the primers, (xvi) the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the average melting temperature of the primers; and (xvii) any combination thereof.

In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have one or more of the following properties: (i)  $\Delta G$  values for possible combinations of that primer with other primers in the library are all equal to or greater than -10 kcal/mol, (ii) the range of melting temperature of the primers is between 1 to 15° C.,

#### US 11,519,035 B2

6

(iii) the length of the target amplicons is between 50 and 100 nucleotides, (iv) the concentration of each primer is less than 20 nM, (v) the length of the annealing step is greater than 5 minutes (such as greater than 10 minutes), (vi) the length of the annealing step is greater than 5 minutes (such greater than 10 minutes), and the concentration of each primer is less than 20 nM, and (vii) any combination thereof. In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have the following properties: (i)  $\Delta G$  values for possible combina- 10 tions of that primer with other primers in the library are all equal to or greater than -10 kcal/mol and (ii) the range of melting temperature of the primers is between 1 to 15° C. In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have the 15 following properties: (i) the length of the target amplicons is between 50 and 100 nucleotides, and (ii) the concentration of each primer is less than 20 nM. In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have the following properties: (i) the 20 length of the target amplicons is between 50 and 100 nucleotides and (ii) the length of the annealing step is greater than 5 minutes (such as greater than 10 minutes). In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have the following 25 properties: (i) the length of the target amplicons is between 50 and 100 nucleotides, (ii) the length of the annealing step is greater than 5 minutes (such greater than 10 minutes), and (iii) the concentration of each primer is less than 20 nM, and (vii) any combination thereof.

In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have one or more of the following properties: (i) the annealing temperature is at least 5° C. (such as at least 6, 8, or 10° C.) greater than the melting temperature (such as the empirically measured 3 or calculated  $T_m$ ) of the primers; (ii) the annealing temperature is between 5 and 15° C., inclusive greater than the melting temperature of the primers; (iii) the annealing temperature is at least 5° C. (such as at least 6, 8, or 10° C.) greater than the highest melting temperature of the primers; 40 (iv) the annealing temperature is at least 5° C. (such as at least 6, 8, or 10° C.) greater than the average melting temperature of the primers, (v) the annealing temperature is between 4 and 15° C. inclusive greater than the average melting temperature of the primers; and (vi) any combination thereof. In various embodiments, at least 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the primers in the library have one or more of the following properties: (i) the annealing temperature is at least 5° C. (such as at least 6, 8, or 10° C.) greater than the melting temperature (such as the empirically 50 measured or calculated  $T_m$ ) of the primers and the length of the annealing step is greater than 5 minutes (such greater than 10 minutes); (ii) the annealing temperature is between 5 and 15° C., inclusive greater than the melting temperature of the primers and the length of the annealing step is greater 55 than 5 minutes (such greater than 10 minutes); (iii) the annealing temperature is at least 5° C. (such as at least 6, 8, or 10° C.) greater than the highest melting temperature of the primers and the length of the annealing step is greater than 5 minutes (such greater than 10 minutes); (iv) the annealing 60 temperature is at least 5° C. (such as at least 6, 8, or 10° C.) greater than the average melting temperature of the primers and the length of the annealing step is greater than 5 minutes (such greater than 10 minutes), (v) the annealing temperature is between 4 and 15° C. inclusive greater than the 65 average melting temperature of the primers and the length of the annealing step is greater than 5 minutes (such greater

68

than 10 minutes); and (vii) any combination thereof. In some embodiments, the guanine-cytosine (GC) content of the primers is between 30% and 80%, inclusive; the range of melting temperatures of the primers is less than 5° C.; and the length of the primers is between 15 to 75 nucleotides, inclusive;

In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include one or more linkages between adjacent nucleotides other than a naturally-occurring phosphodiester linkage. Examples of such linkages include phosphoramide, phosphorothioate, and phosphorodithioate linkages. In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include a thiophosphate (such as a monothiophosphate) between the last 3' nucleotide and the second to last 3' nucleotide. In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include a thiophosphate (such as a monothiophosphate) between the last 2, 3, 4, or 5 nucleotides at the 3' end. In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include a thiophosphate (such as a monothiophosphate) between at least 1, 2, 3, 4, or 5 nucleotides out of the last 10 nucleotides at the 3' end. In some embodiments, such primers are less likely to be cleaved or degraded, such primers may be desirable if a polymerase with proof-reading ability is used (to reduce or prevent the polymerase from removing nucleotides from the primers). In some embodiments, any of the embodiments involving primers with at least one linkage other than a naturally-30 occurring phosphodiester linkage are used with a polymerase having proof-reader activity. In some embodiments, the primers do not contain an enzyme cleavage site (such as a protease cleavage site). In some embodiments, equal to or greater than 1, 10, 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 primers in the library are non-naturally occurring nucleic acids (such nucleic acids with one or more linkages between adjacent nucleotides other than a naturally-occurring phosphodiester linkage).

In some embodiments, the primers have any combination of two or more of the aspects or embodiments disclosed herein.

These primer libraries can be used in any of the methods of the invention.

**Exemplary Primers** 

The primer design methods of the invention have been used to generate several exemplary primer libraries to human target loci. For example, the primer design methods of the invention were used to generate primer libraries. Each of these libraries is composed of three primers per target locus for 1,200; 2,686; or 10,984 different target loci, respectively. The methods of the invention can also be used to generate libraries to non-human target loci.

For an experiment using the 2,686-plex library for multiplex PCR followed by sequencing, the percent of the amplified products that were primer dimers was 11.13%, the median depth of read per target that was amplified was 799.5x coverage, the percent of amplified products that were target amplicons out of the amplified products that were not primer dimers was 93.15% (this is the percent of on target reads when reads for amplified primer dimers are ignored); the number of target loci that were not amplified (failed assay count) was 246; the percent of target loci that were not amplified (failed assay percentage) was 9.16%; the percent of target loci that were amplified was 90.84%; and the total number of reads was 2,522,742. For this primer library, the

## US 11,519,035 B2

6

 $\Delta G$  values for each possible combination of two primers (each possible primer dimer) in the library are all equal to or greater than -3.86 kcal/mol. This -3.86 kcal/mol value was used as a threshold value to select candidate primers that all had a value equal to or greater than (more desirable than) 5 this value from an initial library of candidate primers.

For an experiment using the 10,984-plex library for multiplex PCR followed by sequencing, the percent of the amplified products that were primer dimers was 5.50%, the median depth of read per target that was amplified was 10,286.5× coverage, the percent of amplified products that were target amplicons out of the amplified products that were not primer dimers was 60.16% (this is the percent of on target reads when reads for amplified primer dimers were ignored); the number of target loci that were not amplified (failed assay count) was 3,712; the percent of target loci that were not amplified (failed assay percentage) was 33.79%; the percent of target loci that were amplified was 66.21%; and the total number of reads was 25,372,858.

For an experiment using the 1,200-plex library for multiplex PCR of a sample of only a single cell followed by sequencing, the percent of the amplified products that were primer dimers was 24.13%. This library has primers to human target loci on chromosome 1, chromosome 21, and the X chromosome. For chromosomes 1 and 21, the median depth of read per target that was amplified was 436x coverage; the percent of target loci that were not amplified (failed assay percentage) was 32.69%; and the percent of target loci that were amplified was 67.31%. The total number of reads was 808,106.

The primer design methods of the invention were also used to generate a library for 11,000 different target loci (such as amplifying 10,732 different target human loci using 10,732 different primer pairs). For an experiment using this library for multiplex PCR followed by sequencing, the percent of the amplified products that were primer dimers was 14.75%, the median depth of read per target that was amplified was 72.27× coverage, the percent of the amplified products that were target amplicons was 84.32%; the number of target loci that were not amplified (failed assay count) 40 was 118; the percent of target loci that were not amplified (failed assay percentage) was 1.10%; the percent of target loci that were amplified was 98.9%; and the total number of reads was 6,345,782. For this primer library, the ΔG values for each possible combination of two primers (each possible 45 primer dimer) in the library are all equal to or greater than -4.28 kcal/mol. This -4.28 kcal/mol value was used as a threshold value to select candidate primer that all had a value equal to or greater than (more desirable than) this value from an initial library of candidate primers. For the initial candi- 50 date primers that were used to select primers for this library, the following interaction cost histogram shows the number of candidate primers for each of the following ranges of  $\Delta G$ values. This illustrates how the values for the candidate primers compares to the -4.28 kcal/mol threshold value for 55 the final library.

0 to -0.497 kcal/mol: 88357 -0.497 to -0.993 kcal/mol: 30529 -0.993 to -1.49 kcal/mol: 7862 -1.49 to -1.99 kcal/mol: 2639 -1.99 to -2.48 kcal/mol: 1086 -2.48 to -2.98 kcal/mol: 393 -2.98 to -3.48 kcal/mol: 148 -3.48 to -3.97 kcal/mol: 58 -3.97 to -4.47 kcal/mol: 18 -4.47 to -4.97 kcal/mol: 4 -4.97 to -5.46 kcal/mol: 3 70

-5.46 to -5.96 kcal/mol: 0 -5.96 to -6.46 kcal/mol: 2 -6.46 to -6.95 kcal/mol: 3

The primer design methods of the invention were also used to generate a library for ~14,000 different target loci (such as amplify 13,392 different target human loci with 13,392 different primer pairs). For an experiment using this library for multiplex PCR followed by sequencing, the percent of the amplified products that were primer dimers was 0.56%, the median depth of read per target that was amplified was 69.09x coverage, the percent of the amplified products that were target amplicons was 99.42%; the number of target loci that were not amplified (failed assay count) was 44; the percent of target loci that were not amplified (failed assay percentage) was 0.33%; the percent of target loci that were amplified was 99.67%; and the total number of reads was 7,772,454.

The primer design methods of the invention were also used to generate a library composed of three primers per target locus for 19,488 different target loci. Examples 15, 18, and 19 describe the use of this library. During the PCR amplification and sequencing of a genomic sample, 99.4-99.7% of the sequencing reads mapped to the genome, of those, 99.99% of the reads mapped to target loci. For plasma samples with 10 million sequencing reads, typically at least 19,350 of the 19,488 target loci (99.3%) were amplified and sequenced. For another experiment, the percent of the amplified products that were primer dimers was 1.62%, the median depth of read per target that was amplified was  $30\times$ coverage; the percent of the amplified products that were target amplicons was 98.15%; the number of target loci that were not amplified (failed assay count) was 736; the percent of target loci that were not amplified (failed assay percentage) was 0.56%; the percent of target loci that were amplified was 99.44%; and the total number of reads was 6,476, 975. For this 19,488-plex library, FIG. 34 is a table of the percentage of reads that map to target loci for genomic DNA samples and for samples of a single cell from a cell line for both mother and child samples using this primer library. There was variability in the single cell data which may have resulted from some dead cells being selected, which may have had most of the DNA leaked out. FIG. 35 is an overlay of depth of read for genomic and a single cell sample for different SNPs. FIG. 36 is a table of the percentage of reads that map to target loci for blastoceol fluid and for a single blastocyst cell. The blastoceol fluid produced no mapped reads, possibly due to no DNA being detected. For a single blastocyst, 50-80% of the reads mapped to target loci. FIG. 37 is a graph of reference counts (counts of one allele, such as the "A" allele) divided by total counts for that locus for a single blastocyst cell. For this primer library, the  $\Delta G$  values for each possible combination of two primers (each possible primer dimer) in the library are all equal to or greater than -3.86 kcal/mol. This -3.86 kcal/mol value was used as a threshold value to select candidate primer that all had a value equal to or greater than (more desirable than) this value from an initial library of candidate primers.

The primer design methods of the invention were used to generate a library for ~28,000 different target loci (such as amplifying 27,744 different loci with 27,744 different primer pairs).

For multiplex PCR and sequencing of genomic DNA samples, 99% of the sequencing reads mapped to target loci. The number of different target human loci that were amplified was 23,776.

For an experiment using this library, the percent of the amplified products that were primer dimers was 0.63%, the

#### US 11,519,035 B2

7

median depth of read per target that was amplified was 20x coverage, the percent of the amplified products that were target amplicons was 99.33%; the number of target loci that were not amplified (failed assay count) was 3,968; the percent of target loci that were not amplified (failed assay percentage) was 14.29%; the percent of target loci that were amplified was 85.71%; and the total number of reads was 4,456,636. For a single cell from a cell line, between 2 and 8% of the reads mapped to target loci.

The primer design methods of the invention were used to 10 generate a library for ~9,600 different target loci. As described in Example 10, 7.6 million (97%) of reads mapped to the genome, and 6.3 million (80%) of the reads mapped to the targeted SNPs. The average depth of read was 751, and the median depth of read was 396. As described in 15 Example 9, another experiment produced 3.7 million reads mapping to the genome (94%), and of those, 2.9 million reads (74%) mapped to targeted SNPs with an average depth of read of 344 and a median depth of read of 255.

The primer design methods of the invention were used to 20 generate a library for ~2,400 different target loci. As described in Example 12, when four portions were each amplified with ~2,400 primers, 4.5 million reads mapped to targeted SNPs, the average depth of read was 535 and the median depth of read was 412.

If desired, any of the results may be improved by increasing the number of cells or the amount of nucleic acid template used for the analysis or by optimizing the conditions. For example, if results from single cell samples are not as good as desired for a particular application, a sample with more cells or more nucleic acids may be used instead (such as to decrease the percentage of primer dimers, increase the percentage of target amplicons, or increase the percentage of target loci that are amplified). Samples with more nucleic acids have more template molecules for the primers to bind dimers binding each other and forming primer dimers.

These primer libraries or primer pools can be used in any of the PCR methods of the invention. In some embodiments, primers from any of the primer pools are used in combination with a universal primer to amplify the target loci. In some embodiments, multiple rounds of PCR are performed in which each round of PCR uses primers from one of the primer pools and a universal primer. In some embodiments, primers from two of the primer pools are used to amplify the target loci. In some embodiments, multiple rounds of PCR are performed. In some embodiments, primers from pools C and B are used for the first round of PCR and then primers from pools A and C are used for the second of PCR. In some embodiments, primers from pools C and B are used for the first round of PCR and then primers from pools A and B are used for the second of PCR.

In some embodiments, a region that is not specific for a target locus (such as a tag, bar code, or universal binding site) is added to one or more primers of the invention. In 55 various embodiments, the nonspecific region is added to the 5' end of the primer, to the 3' end of the primer, or to an internal region of the primer. In some embodiments, the primers are fragments (such as fragments of at least 10, 20, 30, 40, 50 or more contiguous nucleotides that are less than 60 full-length).

In some embodiments, the invention provides a library of primers that includes at least 10; 20; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 20,000; 25,000; 30,000; 40,000; or 50,000 different primers. In some 65 embodiments, the invention provides a library of primers that includes at least 10; 20; 50; 75; 100; 300; 500; 750;

72

1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 20,000; 25,000; 30,000; 40,000; or 50,000 different primers.

Percent identity in reference to nucleic acid sequences refers to the degree of sequence identity between nucleic acid sequences. Percent identity can be determined in various ways that are within the skill in the art, for instance, using publicly available computer software with the default parameters such as Smith Waterman Alignment (Smith and Waterman J. Mol. Biol. 147:195-7, 1981); "BestFit" (Smith and Waterman, Advances in Applied Mathematics, 482-489, 1981); Basic Local Alignment Search Tool (BLAST, Altschul, S. F., W. Gish, et al., J. Mol. Biol. 215: 403-410, 1990; available through the U.S. government's National Center for Biotechnology Information web site at the world wide web at ncbi.nlm.nih.gov), BLAST-2, BLAST-N, WU-BLAST, WU-BLAST-2, ENTREZ (available through the National Center for Biotechnology Information), CLUSTALW, CLUSTAL Omega, or Megalign (DNASTAR, Inc.

Madison, Wis.) software. In addition, those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the length of the sequences being compared. In general, the length of comparison will generally be at least 20, 30, 40, 45, 50, or more nucleotides.

In some embodiments, percent identity is calculated by determining the number of matched positions in aligned nucleic acid sequences, dividing the number of matched positions by the total number of aligned nucleotides, and multiplying by 100. A matched position refers to a position in which identical nucleotides occur at the same position in aligned nucleic acid sequences. The percent identity over a particular length is determined by counting the number of matched positions over that length and dividing that number by the length followed by multiplying the resulting value by 100. For example, if (i) a 500-nucleotide nucleic acid target sequence is compared to a subject nucleic acid sequence, (ii) an alignment program presents 200 nucleotides from the target sequence aligned with a region of the subject sequence where the first and last nucleotides of that 200-nucleotide region are matches, and (iii) the number of matches over those 200 aligned nucleotides is 180, then the 500-nucleotide nucleic acid target sequence contains a length of 200 and a sequence identity over that length of 90% (i.e., 180, 200×100=90).

Hybridization conditions resulting in a particular degree of stringency will vary depending upon the nature of the hybridization method and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (such as the Na<sup>+</sup> concentration) of the hybridization buffer will determine the stringency of hybridization. Calculations regarding hybridization conditions for attaining particular degrees of stringency are discussed in Sambrook et al., (1989) Molecular Cloning, second edition, Cold Spring Harbor Laboratory, Plainview, N.Y. (chapters 9 and 11); Nucleic Acid Hybridization, A Practical Approach, Ed. Hames, B. D. and Higgins, S. J., IRL Press, 1985; Ausubel et al. Current Protocols in Molecular Biology, Wiley, New York 1994; and U.S. Pat. No. 8,357,488, filed May 16, 2008. In some embodiments, very high stringency hybridization conditions includes an overnight incubation at 42° C. in a solution comprising 50% formamide, 5×SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5×Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1×SSC at about 65° C. The following is an exemplary set of hybridization conditions and is not limiting:

#### US 11,519,035 B2

73

Very High Stringency (Detects Sequences that Share at Least 90% Identity)

Hybridization: 5×SSC at 65° C. for 16 hours

Wash twice: 2×SSC at room temperature for 15 minutes each

Wash twice: 0.5×SSC at 65° C. for 20 minutes each High Stringency (Detects Sequences that Share at Least 80% Identity)

Hybridization:  $5\times-6\times SSC$  at  $65^{\circ}$  C. $-70^{\circ}$  C. for 16-20 hours

Wash twice: 2×SSC at room temperature for 5-20 minutes each

Wash twice: 1×SSC at 55° C.-70° C. for 30 minutes each Low Stringency (Detects Sequences that Share at Least 50% Identity)

Hybridization:  $6\times SSC$  at room temperature to  $55^{\circ}$  C. for 16-20 hours

Wash at least twice: 2x-3xSSC at room temperature to 55° C. for 20-30 minutes each

These primers can be used in any of the primer libraries 20 or methods of the invention.

Exemplary Primer Libraries for Detection of Recombination In some embodiments, primers in the primer library are designed to determine whether or not recombination occurred at one or more known recombination hotspots 25 (such as crossovers between homologous human chromosomes). Knowing what crossovers occurred between chromosomes allows more accurate phased genetic data to be determined for an individual. Recombination hotspots are local regions of chromosomes in which recombination 30 events tend to be concentrated. Often they are flanked by "coldspots," regions of lower than average frequency of recombination. Recombination hotspots tend to share a similar morphology and are approximately 1 to 2 kb in length. The hotspot distribution is positively correlated with 3: GC content and repetitive element distribution. A partially degenerated 13-mer motif CCNCCNTNNCCNC plays a role in some hotspot activity. It has been shown that the zinc finger protein called PRDM9 binds to this motif and initiates recombination at its location. The average distance between 40 the centers of recombination hot spots is reported to be 80 kb. In some embodiments, the distance between the centers of recombination hot spots ranges between 3 kb to 100 kb. Public databases include a large number of known human recombination hotspots, such as the HUMHOT and Inter- 45 national HapMap Project databases (see, for example, Nishant et al., "HUMHOT: a database of human meiotic recombination hot spots," Nucleic Acids Research, 34: D25-D28, 2006, Database issue; Mackiewicz et al., "Distribution of Recombination Hotspots in the Human Genome—A 50 Comparison of Computer Simulations with Real Data" e65272, doi:10.1371/journal. **PLoS** ONE 8(6): pone.0065272; and the world wide web at hapmap.ncbi.nlm-.nih.gov/downloads/index.html.en, which are each hereby incorporated by reference in its entirety).

In some embodiments, primers in the primer library are clustered at or near recombination hotspots (such as known human recombination hotspots). In some embodiments, the corresponding amplicons are used to determine the sequence within or near a recombination hotspot to determine whether or not recombination occurred at that particular hotspot (such as whether the sequence of the amplicon is the sequence expected if a recombination had occurred or the sequence expected if a recombination had not occurred). In some embodiments, primers are designed to amplify part or all of a recombination hotspot (and optionally sequence flanking a recombination hotspot). In some embodiments,

74

long read sequencing (such as sequencing using the Moleculo Technology developed by Illumina to sequence up to 10 kb) or paired end sequencing is used to sequence part or all of a recombination hotspot. Knowledge of whether or not a recombination event occurred can be used to determine which haplotype blocks flank the hotspot. If desired, the presence of particular haplotype blocks can be confirmed using primers specific to regions within the haplotype blocks. In some embodiments, it is assumed there are no crossovers between known recombination hotspots. In some embodiments, primers in the primer library are clustered at or near the ends of chromosomes. For example, such primers can be used to determine whether or not a particular arm or section at the end of a chromosome is present. In some embodiments, primers in the primer library are clustered at or near recombination hotspots and at or near the ends of

In some embodiments, the primer library includes one or more primers (such as at least 5; 10; 50; 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; or 50,000 different primers or different primer pairs) that are specific for a recombination hotspot (such as a known human recombination hotspot) and/or are specific for a region near a recombination hotspot (such as within 10, 8, 5, 3, 2, 1, or 0.5 kb of the 5' or 3' end of a recombination hotspot). In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for the same recombination hotspot, or are specific for the same recombination hotspot or a region near the recombination hotspot. In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for a region between recombination hotspots (such as a region unlikely to have undergone recombination); these primers can be used to confirm the presence of haplotype blocks (such as those that would be expected depending on whether or not recombination has occurred). In some embodiments, at least 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the primers in the primer library are specific for a recombination hotspot and/or are specific for a region near a recombination hotspot (such as within 10, 8, 5, 3, 2, 1, or 0.5 kb of the 5' or 3' end of the recombination hotspot). In some embodiments, the primer library is used to determine whether or not recombination has occurred at greater than or equal to 5; 10; 50; 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; or 50,000 different recombination hotspots (such as known human recombination hotspots). In some embodiments, the regions targeted by primers to a recombination hotspot or nearby region are approximately evenly spread out along that portion of the genome. In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for the a region at or near the end of a chromosome (such as a region within 20, 10, 5, 1, 0.5, 0.1, 0.01, or 0.001 mb from the end of a chromosome). In some embodiments, at least 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the primers in the primer library are specific for the a region at or near the end of a chromosome (such as a region within 20, 10, 5, 1, 0.5, 0.1, 0.01, or 0.001 mb from the end of a chromosome). In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for a region within a potential microdeletion in a chromosome. In some embodiments, at least 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the primers in the primer library are specific for a region within a potential microdeletion in a chromosome. In some embodiments, at least 10, 20, 30, 40,  $50,\,60,\,70,\,80,$  or 90% of the primers in the primer library are specific for a recombination hotspot, a region near a

#### US 11,519,035 B2

7:

recombination hotspot, a region at or near the end of a chromosome, or a region within a potential microdeletion in a chromosome. In some embodiments, the primers have one or more of the properties described herein. Other embodiments are disclosed in U.S. Ser. No. 61/987,407, filed May 5 1, 2014 and 62/066,514, filed Oct. 21, 2014. Exemplary Kits

In one aspect, the invention features a kit, such as a kit for amplifying target loci in a nucleic acid sample for detecting deletions and/or duplications of chromosome segments or 10 entire chromosomes using any of the methods described herein). In some embodiments, the kit can include any of the primer libraries of the invention. In an embodiment, the kit comprises a plurality of inner forward primers and optionally a plurality of inner reverse primers, and optionally outer 15 forward primers and outer reverse primers, where each of the primers is designed to hybridize to the region of DNA immediately upstream and/or downstream from one of the target sites (e.g., polymorphic sites) on the target chromosome(s) or chromosome segment(s), and optionally addi- 20 tional chromosomes or chromosome segments. In some embodiments, the kit includes instructions for using the primer library to amplify the target loci, such as for detecting one or more deletions and/or duplications of one or more chromosome segments or entire chromosomes using any of 25 the methods described herein.

In certain embodiments, kits of the invention provide primer pairs for detecting chromosomal aneuploidy and CNV determination, such as primer pairs for massively multiplex reactions for detecting chromosomal aneuploidy 30 such as CNV (CoNVERGe) (Copy Number Variant Events Revealed Genotypically) and/or SNVs. In these embodiments, the kits can include between at least 100, 200, 250, 300, 500, 1000, 2000, 2500, 3000, 5000, 10,000, 20,000, 25,000, 28,000, 50,000, or 75,000 and at most 200, 250, 300, 500, 1000, 2000, 2500, 3000, 5000, 10,000, 20,000, 25,000, 28,000, 50,000, 75,000, or 100,000 primer pairs that are shipped together. The primer pairs can be contained in a single vessel, such as a single tube or box, or multiple tubes or boxes. In certain embodiments, the primer pairs are 40 pre-qualified by a commercial provider and sold together, and in other embodiments, a customer selects custom gene targets and/or primers and a commercial provider makes and ships the primer pool to the customer neither in one tube or a plurality of tubes. In certain exemplary embodiments, the 45 kits include primers for detecting both CNVs and SNVs, especially CNVs and SNVs known to be correlated to at least one type of cancer.

Kits for circulating DNA detection according to some embodiments of the present invention, include standards 50 and/or controls for circulating DNA detection. For example, in certain embodiments, the standards and/or controls are sold and optionally shipped and packaged together with primers used to perform the amplification reactions provided herein, such as primers for performing CoNVERGe. In 55 certain embodiments, the controls include polynucleotides such as DNA, including isolated genomic DNA that exhibits one or more chromosomal aneuploidies such as CNV and/or includes one or more SNVs. In certain embodiments, the standards and/or controls are called PlasmArt standards and 60 include polynucleotides having sequence identity to regions of the genome known to exhibit CNV, especially in certain inherited diseases, and in certain disease states such as cancer, as well as a size distribution that reflects that of cfDNA fragments naturally found in plasma. Exemplary 65 methods for making PlasmArt standards are provided in the examples herein. In general, genomic DNA from a source

76

known to include a chromosomal aneuoploidy is isolated, fragmented, purified and size selected.

Accordingly, artificial cfDNA polynucleotide standards and/or controls can be made by spiking isolated polynucleotide samples prepared as summarized above, into DNA samples known not to exhibit a chromosomal aneuploidy and/or SNVs, at concentrations similar to those observed for cfDNA in vivo, such as between, for example, 0.01% and 20%, 0.1 and 15%, or 0.4 and 10% of DNA in that fluid. These standards/controls can be used as controls for assay design, characterization, development, and/or validation, and as quality control standards during testing, such as cancer testing performed in a CLIA lab and/or as standards included in research use only or diagnostic test kits.

**Exemplary Amplicons** 

In one aspect, the invention provides a composition comprising at least 100 different amplicons (e.g., at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical amplicons) in solution in one reaction volume. In some embodiments, the amplicons are produced from the simultaneous PCR amplification of at least 100 different target loci (e.g., at least 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical target loci) using at least 100 different primers or primer pairs (e.g., at least 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers or primer pairs) in one reaction volume. In some embodiments, (i) less than 60% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (ii) less than 40% of the amplified products are primer dimers and at least 60% of the amplified products are target amplicons, (iii) less than 20% of the amplified products are primer dimers and at least 80% of the amplified products are target amplicons, (iv) less than 10% of the amplified products are primer dimers and at least 90% of the amplified products are target amplicons, or (v) less than 5% of the amplified products are primer dimers and at least 95% of the amplified products are target amplicons. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the range of the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, the composition includes at least 1,000 different amplicons in solution in one reaction volume; wherein the amplicons are produced from the simultaneous PCR amplification of at least 1,000 different target human loci using at least 1,000 different primers in one reaction volume; wherein (i) less than 20% of the amplicons are primer dimers, and (ii) at least 80% of the amplicons comprise one of the target human loci and are between 50 and 100 nucleotides in length, inclusive. In some embodiments, the composition consists essentially of, or consists of one or more of the following: amplicons, primers (such as any of the primers disclosed herein), free nucleotide(s), non-human or non-naturally occurring enzyme(s), buffer(s), or any combination thereof.

In some embodiments, a large percentage or substantially all of the primers used for the multiplex PCR method are consumed during the PCR reaction or are removed from the reaction volume after the PCR amplification. In some

#### US 11,519,035 B2

7

embodiments, at least 80, 90, 92, 94, 96, 98, 99, or 100% of the primer molecules are extended to form amplified products. In some embodiments, for at least 80, 90, 92, 94, 96, 98, 99, or 100% of target loci, at least 80, 90, 92, 94, 96, 98, 99, or 100% of the primer molecules to that target loci are extended to form amplified products. In some embodiments, multiple cycles are performed until all or substantially all of the primers are consumed. If desired, a higher percentage of the primers can be consumed by decreasing the initial primer concentration and/or increasing the number of PCR cycles that are performed. In some embodiments, at least 80, 90, 95, 96, 97, 98, 99, or 100% of the nucleic acids in the composition are amplicons (instead of unextended dimers).

In one aspect, the invention provides a composition comprising at least 100 different primers or primer pairs 15 (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical primers or primer pairs) and at least 100 different amplicons (e.g., at least 300, 500; 750; 1,000; 20 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical amplicons) in solution in one reaction volume. In some embodiments, the amplicons are produced from the simultaneous PCR amplification of at least 100 25 different target loci (e.g., at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical target loci) using the primers or primer pairs in one reaction volume. In some 30 embodiments, (i) less than 60% of the amplified products are primer dimers and at least 40% of the amplified products are target amplicons, (ii) less than 40% of the amplified products are primer dimers and at least 60% of the amplified products are target amplicons, (iii) less than 20% of the amplified products are primer dimers and at least 80% of the amplified products are target amplicons, (iv) less than 10% of the amplified products are primer dimers and at least 90% of the amplified products are target amplicons, or (v) less than 5% of the amplified products are primer dimers and at least 95% 40 of the amplified products are target amplicons. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less 45 than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the range of the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, the composition comprising at least 1,000 50 different primers and at least 1,000 different amplicons in solution in one reaction volume; wherein the amplicons are produced from the simultaneous PCR amplification of at least 1,000 different target human loci with the primers in one reaction volume; wherein (i) less than 20% of the 55 amplicons are primer dimers, and (ii) at least 80% of the amplicons comprise one of the target loci and are between 50 and 100 nucleotides in length, inclusive. In some embodiments, the composition consists essentially of, or consists of one or more of the following: amplicons, primers (such as 60 any of the primers disclosed herein), free nucleotide(s), non-human or non-naturally occurring enzyme(s), buffer(s), or any combination thereof.

In some embodiments, the amplification of different target loci is substantially uniform. In some embodiments, target 65 loci (such as nonpolymorphic target loci or polymorphic target loci that are amplified regardless of what allele is

78

present at the polymorphic site) that were present in the same amount (or substantially the same amount) in the initial unamplified sample are also present in substantially the same amount in the PCR-amplified products. In some embodiments, for at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different targets, the target loci that is amplified the most out of these targets (which can be all of the targets or a subset of the targets for a library) is amplified less than 2,000; 1,500; 1,000; 500, 400, 300, 200, 100%, 50, 20, 10, 5, or 2% more than the target loci that is amplified the least out of these targets. In some embodiments, for at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target amplicons, the target amplicon in greatest abundance out of these target amplicons (which can be all of the target amplicons or a subset of the target amplicons produced by a library) is present in an amount that is less than 2,000; 1,500; 1,000; 500, 400, 300, 200, 100%, 50, 20, 10, 5, or 2% more than the target amplicon in least abundance out of these target amplicons. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target amplicons are present in an amount that is at least 5, 10, 15, 20, 40, 50, 60, 70, 80, or 90% of the amount of the target amplicon in greatest abundance. In some embodiments, for at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different amplicons that are produced by multiplex PCR and then sequenced, at least 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target amplicons produce a number of sequencing reads within 20, 30, 50, or 80% above or below the mean number of sequences for target amplicons. If desired, the uniformity can be increased by using primers with more similar primer lengths, target amplicon lengths, GC contents, melting temperatures, or any combination thereof. In some embodiments, the uniformity can be increased by using TMAC in the reaction volume during amplification. In some embodiments, having most or all the primers consumed in the PCR reaction increases the uniformity of amplification.

If desired, the uniformity in DOR can be measured using standard methods such as depth of read slope (DOR slope), normalized median depth of read (nmDOR), or breadth of read (BOR). DOR slope represents the slope of the line in the linear portion of a list of loci sorted in descending DOR order. Closer to zero is better, as it represents a flat line. In some embodiments, the uniformity in DOR can be measured using the Percent of reads in the 90'-95<sup>th</sup> Percentile. For this measurement, the loci are sorted in descending DOR order. In the ideal DOR distribution, the 90-95 percentile should contain 5% of reads. The reads of all loci between the 90<sup>th</sup> Percentile and 95<sup>th</sup> percentile are counted and divided by the total reads for all loci. In one experiment, the DOR slope versus percent of reads in the 90<sup>th</sup>-95<sup>th</sup> percentile for all samples had an R<sup>2</sup>=0.81.

In some embodiments, the magnitude of the DOR slope is less than 0.005, 0.001, 0.0005, 0.0001, 0.00005, 0.0001, 0.00005, 0.00011, 0.00005, or 0.000001. In some embodiments, the magnitude of the DOR slope is between 0 and 0.005, such as 0.000001 to 0.005, such as between 0.000005 to 0.00001, 0.00001 to 0.00005, 0.00005 to 0.0001, 0.0001 to 0.0005, 0.0005 to 0.001, or 0.001 to 0.005, inclusive. In some embodiments, the percent of reads in the 90<sup>th</sup>-95<sup>th</sup> percentile

#### US 11,519,035 B2

7

is between 0.2 and 9%, such as between 1 to 8%, 2 to 7%, 0.2 to 1.0%, 1 to 2%, 2 to 3%, 2 to 4%, 3 to 4%, 4 to 5%, 5 to 6%, or 6 to 8%, or 7 to 9& inclusive. In some embodiments, the invention features a composition comprising at least 100 different amplicons (e.g., at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical amplicons) with the magnitude of the DOR slope in any of these ranges or with a percent of reads in the 90<sup>th</sup>-95<sup>th</sup> percentile in any of 10 these ranges. In some embodiments, the amplification method produces at least 100 different amplicons (e.g., at least 300, 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 non-identical ampli- 15 cons) with the magnitude of the DOR slope in any of these ranges or with a percent of reads in the 90th-95th percentile in any of these ranges.

# Exemplary Multiplex PCR Methods

In one aspect, the invention features methods of ampli- 20 fying target loci in a nucleic acid sample that involve (i) contacting the nucleic acid sample with a library of primers that simultaneously hybridize to least 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different 25 target loci to produce a reaction mixture; and (ii) subjecting the reaction mixture to primer extension reaction conditions (such as PCR conditions) to produce amplified products that include target amplicons. In some embodiments, the method also includes determining the presence or absence of at least 30 one target amplicon (such as at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target amplicons). In some embodiments, the method also includes determining the sequence of at least one target amplicon (such as at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target 3: amplicons). In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified. In some embodiments, at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 40 50,000; 75,000; or 100,000 different target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 200, 300, or 400-fold. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.5, or 100% of the target loci are amplified at least 5, 10, 20, 40, 50, 60, 80, 100, 120, 150, 45 200, 300, or 400-fold. In various embodiments, less than 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.05% of the amplified products are primer dimers.

In some embodiments, the method involves multiplex PCR and sequencing (such as high throughput sequencing). 50

In various embodiments, long annealing times and/or low primer concentrations are used. In various embodiments, the length of the annealing step is greater than 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes. In various embodiments, the length of the annealing step (per PCR 55 cycle) is between 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive. In various embodiments, the length of the annealing step is greater than 5 minutes (such greater than 10, or 15 minutes), and the concentration of each primer is less than 20 nM. In various 60 embodiments, the length of the annealing step is greater than 5 minutes (such greater than 10, or 15 minutes), and the concentration of each primer is between 1 to 20 nM, or 1 to 10 nM, inclusive. In various embodiments, the length of the annealing step is greater than 20 minutes (such as greater 65 than 30, 45, 60, or 90 minutes), and the concentration of each primer is less than 1 nM.

80

At high level of multiplexing, the solution may become viscous due to the large amount of primers in solution. If the solution is too viscous, one can reduce the primer concentration to an amount that is still sufficient for the primers to bind the template DNA. In various embodiments, less than 60,000 different primers are used and the concentration of each primer is less than 20 nM, such as less than 10 nM or between 1 and 10 nM, inclusive. In various embodiments, more than 60,000 different primers (such as between 60,000 and 120,000 different primers) are used and the concentration of each primer is less than 10 nM, such as less than 5 nM or between 1 and 10 nM, inclusive.

It was discovered that the annealing temperature can optionally be higher than the melting temperatures of some or all of the primers (in contrast to other methods that use an annealing temperature below the melting temperatures of the primers) (Example 25). The melting temperature  $(T_m)$  is the temperature at which one-half (50%) of a DNA duplex of an oligonucleotide (such as a primer) and its perfect complement dissociates and becomes single strand DNA. The annealing temperature (TA) is the temperature one runs the PCR protocol at. For prior methods, it is usually 5° C. below the lowest  $T_m$  of the primers used, thus close to all possible duplexes are formed (such that essentially all the primer molecules bind the template nucleic acid). While this is highly efficient, at lower temperatures there are more unspecific reactions bound to occur. One consequence of having too low a TA is that primers may anneal to sequences other than the true target, as internal single-base mismatches or partial annealing may be tolerated. In some embodiments of the present inventions, the  $T_A$  is higher than  $(T_m)$ , where at a given moment only a small fraction of the targets have a primer annealed (such as only ~1-5%). If these get extended, they are removed from the equilibrium of annealing and dissociating primers and target (as extension increases  $T_m$ quickly to above 70 C), and a new 1-5% of targets has primers. Thus, by giving the reaction long time for annealing, one can get ~100% of the targets copied per cycle. Thus, the most stable molecule pairs (those with perfect DNA pairing between the primer and the template DNA) are preferentially extended to produce the correct target amplicons. For example, the same experiment was performed with 57° C. as the annealing temperature and with 63° C. as the annealing temperature with primers that had a melting temperature below 63° C. When the annealing temperature was 57° C., the percent of mapped reads for the amplified PCR products was as low as 50% (with 50% of the amplified products being primer-dimer). When the annealing temperature was 63° C., the percentage of amplified products that were primer dimer dropped to ~2%.

In various embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers. In some embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers, and the length of the annealing step (per PCR cycle) is greater than 1, 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes.

#### US 11,519,035 B2

8

In various embodiments, the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers. In various embodiments, the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 10 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 15 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers, and the length of the annealing step (per PCR cycle) is between 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive.

In some embodiments, the annealing temperature is at 20 least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the highest melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers. In some embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the highest 25 melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers, and the length of the annealing step (per PCR cycle) is greater than 1, 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes

In some embodiments, the annealing temperature is 30 between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the highest melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers. In some embodiments, the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the highest melting temperature (such as the empirically measured or calculated  $T_m$ ) of the primers, and the length of the annealing step (per PCR cycle) is between 40 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive.

In some embodiments, the annealing temperature is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the average melting temperature (such as the empiri- 45 cally measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers. In some embodiments, the annealing temperature is at least 50 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 15° C. greater than the average melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 55 50,000; 75,000; 100,000; or all of the non-identical primers, and the length of the annealing step (per PCR cycle) is greater than 1, 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes.

In some embodiments, the annealing temperature is 60 between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the average melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 65 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of

82

the non-identical primers. In some embodiments, the annealing temperature is between 1 and 15° C. (such as between 1 to 10, 1 to 5, 1 to 3, 3 to 5, 5 to 10, 5 to 8, 8 to 10, 10 to 12, or 12 to 15° C., inclusive) greater than the average melting temperature (such as the empirically measured or calculated  $T_m$ ) of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; 100,000; or all of the non-identical primers, and the length of the annealing step (per PCR cycle) is between 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive.

In some embodiments, the annealing temperature is between 50 to 70° C., such as between 55 to 60, 60 to 65, or 65 to 70° C., inclusive. In some embodiments, the annealing temperature is between 50 to 70° C., such as between 55 to 60, 60 to 65, or 65 to 70° C., inclusive, and either (i) the length of the annealing step (per PCR cycle) is greater than 3, 5, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, or 180 minutes or (ii) the length of the annealing step (per PCR cycle) is between 5 and 180 minutes, such as 5 to 60, 10 to 60, 5 to 30, or 10 to 30 minutes, inclusive.

In some embodiments, one or more of the following conditions are used for empirical measurement of  $T_m$  or are assumed for calculation of  $T_m$ : temperature: of 60.0° C., primer concentration of 100 nM, and/or salt concentration of 100 mM. In some embodiments, other conditions are used, such as the conditions that will be used for multiplex PCR with the library. In some embodiments, 100 mM KCl, 50 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 mM MgCl<sub>2</sub>, 7.5 nM of each primer, and 50 mM TMAC, at pH 8.1 is used. In some embodiments, the  $T_m$  is calculated using the Primer3 program (libprimer3 release 2.2.3) using the built-in SantaLucia parameters (the world wide web at primer3.sourceforge.net, which is hereby incorporated by reference in its entirety). For example, the  $T_m$  values may be calculated using the method in Example 25. In some embodiments, the calculated melting temperature for a primer is the temperature at which half of the primers molecules are expected to be annealed. As discussed above, even at a temperature higher than the calculated melting temperature, a percentage of primers will be annealed, and therefore PCR extension is possible. In some embodiments, the empirically measured  $T_m$  (the actual  $T_m$ ) is determined by using a thermostatted cell in a UV spectrophotometer. In some embodiments, temperature is plotted vs. absorbance, generating an S-shaped curve with two plateaus. The absorbance reading halfway between the plateaus corresponds to  $T_m$ .

In some embodiments, the absorbance at 260 nm is measured as a function of temperature on an ultrospec 2100 pr UV/visible spectrophotometer (Amershambiosciences) (see, e.g., Takiya et al., "An empirical approach for thermal stability (T<sub>m</sub>) prediction of PNA/DNA duplexes," Nucleic Acids Symp Ser (Oxf); (48):131-2, 2004, which is hereby incorporated by reference in its entirety). In some embodiments, absorbance at 260 nm is measured by decreasing the temperature in steps of 2° C. per minute from 95 to 20° C. In some embodiments, a primer and its perfect complement (such as 2 uM of each paired oligomer) are mixed and then annealing is performed by heating the sample to 95° C., keeping it there for 5 minutes, followed by cooling to room temperature during 30 minutes, and keeping the samples at 95° C. for at least 60 minutes.

In some embodiments, melting temperature is determined by analyzing the data using SWIFT  $T_m$  software. In some embodiments of any of the methods of the invention, the method includes empirically measuring or calculating (such

#### US 11,519,035 B2

8

as calculating with a computer) the melting temperature for at least 50, 80, 90, 92, 94, 96, 98, 99, or 100% of the primers in the library either before or after the primers are used for PCR amplification of target loci.

In some embodiments, the library comprises a microarray. 5 In some embodiments, the library does not comprise a microarray.

In some embodiments, most or all of the primers are extended to form amplified products. Having all the primers consumed in the PCR reaction increases the uniformity of 10 amplification of the different target loci since the same or similar number of primer molecules are converted to target amplicons for each target loci. In some embodiment, at least 80, 90, 92, 94, 96, 98, 99, or 100% of the primer molecules are extended to form amplified products. In some embodi- 15 ments, for at least 80, 90, 92, 94, 96, 98, 99, or 100% of target loci, at least 80, 90, 92, 94, 96, 98, 99, or 100% of the primer molecules to that target loci are extended to form amplified products. In some embodiments, multiple cycles are performed until this percentage of the primers are 20 consumed. In some embodiments, multiple cycles are performed until all or substantially all of the primers are consumed. If desired, a higher percentage of the primers can be consumed by decreasing the initial primer concentration and/or increasing the number of PCR cycles that are per- 25 formed.

In some embodiments, the PCR methods may be performed with microliter reaction volumes, for which it can be harder to achieve specific PCR amplification (due to the lower local concentration of the template nucleic acids) 30 compared to nanoliter or picoliter reaction volumes used in microfluidics applications. In some embodiments, the reaction volume is between 1 and 60 uL, such as between 5 and 50 uL, 10 and 50 uL, 10 and 20 uL, 20 and 30 uL, 30 and 40 uL, or 40 to 50 uL, inclusive.

In an embodiment, a method disclosed herein uses highly efficient highly multiplexed targeted PCR to amplify DNA followed by high throughput sequencing to determine the allele frequencies at each target locus. The ability to multiplex more than about 50 or 100 PCR primers in one reaction 40 volume in a way that most of the resulting sequence reads map to targeted loci is novel and non-obvious. One technique that allows highly multiplexed targeted PCR to perform in a highly efficient manner involves designing primers that are unlikely to hybridize with one another. The PCR 45 probes, typically referred to as primers, are selected by creating a thermodynamic model of potentially adverse interactions between at least 300; at least 500; at least 750; at least 1,000; at least 2,000; at least 5,000; at least 7,500; at least 10,000; at least 20,000; at least 25,000; at least 30,000; 50 at least 40,000; at least 50,000; at least 75,000; or at least 100,000 potential primer pairs, or unintended interactions between primers and sample DNA, and then using the model to eliminate designs that are incompatible with other the designs in the pool. Another technique that allows highly multiplexed targeted PCR to perform in a highly efficient manner is using a partial or full nesting approach to the targeted PCR. Using one or a combination of these approaches allows multiplexing of at least 300, at least 800, at least 1,200, at least 4,000 or at least 10,000 primers in a 60 single pool with the resulting amplified DNA comprising a majority of DNA molecules that, when sequenced, will map to targeted loci. Using one or a combination of these approaches allows multiplexing of a large number of primers in a single pool with the resulting amplified DNA 65 comprising greater than 50%, greater than 60%, greater than 67%, greater than 80%, greater than 90%, greater than 95%,

84

greater than 96%, greater than 97%, greater than 98%, greater than 99%, or greater than 99.5% DNA molecules that map to targeted loci.

In some embodiments the detection of the target genetic material may be done in a multiplexed fashion. The number of genetic target sequences that may be run in parallel can range from one to ten, ten to one hundred, one hundred to one thousand, one thousand to ten thousand, ten thousand to one hundred thousand, one hundred thousand to one million, or one million to ten million. Prior attempts to multiplex more than 100 primers per pool have resulted in significant problems with unwanted side reactions such as primer-dimer formation.

Targeted PCR

In some embodiments, PCR can be used to target specific locations of the genome. In plasma samples, the original DNA is highly fragmented (typically less than 500 bp, with an average length less than 200 bp). In PCR, both forward and reverse primers anneal to the same fragment to enable amplification. Therefore, if the fragments are short, the PCR assays must amplify relatively short regions as well. Like MIPS, if the polymorphic positions are too close the polymerase binding site, it could result in biases in the amplification from different alleles.

Currently, PCR primers that target polymorphic regions, such as those containing SNPs, are typically designed such that the 3' end of the primer will hybridize to the base immediately adjacent to the polymorphic base or bases. In an embodiment of the present disclosure, the 3' ends of both the forward and reverse PCR primers are designed to hybridize to bases that are one or a few positions away from the variant positions (polymorphic sites) of the targeted allele.

The number of bases between the polymorphic site (SNP or otherwise) and the base to which the 3' end of the primer is designed to hybridize may be one base, it may be two bases, it may be three bases, it may be four bases, it may be five bases, it may be seven to ten bases, it may be eleven to fifteen bases, or it may be sixteen to twenty bases. The forward and reverse primers may be designed to hybridize a different number of bases away from the polymorphic site.

PCR assay can be generated in large numbers, however, the interactions between different PCR assays makes it difficult to multiplex them beyond about one hundred assays. Various complex molecular approaches can be used to increase the level of multiplexing, but it may still be limited to fewer than 100, perhaps 200, or possibly 500 assays per reaction. Samples with large quantities of DNA can be split among multiple sub-reactions and then recombined before sequencing. For samples where either the overall sample or some subpopulation of DNA molecules is limited, splitting the sample would introduce statistical noise. In an embodiment, a small or limited quantity of DNA may refer to an amount below 10 µg, between 10 and 100 µg, between 100  $\mu g$  and 1 ng, between 1 and 10 ng, or between 10 and 100 ng. Note that while this method is particularly useful on small amounts of DNA where other methods that involve splitting into multiple pools can cause significant problems related to introduced stochastic noise, this method still provides the benefit of minimizing bias when it is run on samples of any quantity of DNA. In these situations a universal pre-amplification step may be used to increase the overall sample quantity. Ideally, this pre-amplification step should not appreciably alter the allelic distributions.

In an embodiment, a method of the present disclosure can generate PCR products that are specific to a large number of

#### US 11,519,035 B2

8:

targeted loci, specifically 1,000 to 5,000 loci, 5,000 to 10,000 loci or more than 10,000 loci, for genotyping by sequencing or some other genotyping method, from limited samples such as single cells or DNA from body fluids. Currently, performing multiplex PCR reactions of more than 5 5 to 10 targets presents a major challenge and is often hindered by primer side products, such as primer dimers, and other artifacts. When detecting target sequences using microarrays with hybridization probes, primer dimers and other artifacts may be ignored, as these are not detected. However, when using sequencing as a method of detection, the vast majority of the sequencing reads would sequence such artifacts and not the desired target sequences in a sample. Methods described in the prior art used to multiplex more than 50 or 100 reactions in one reaction volume followed by sequencing will typically result in more than 20%, and often more than 50%, in many cases more than 80% and in some cases more than 90% off-target sequence reads

In general, to perform targeted sequencing of multiple (n) targets of a sample (greater than 50, greater than 100, greater than 500, or greater than 1,000), one can split the sample into a number of parallel reactions that amplify one individual target. This has been performed in PCR multiwell plates or can be done in commercial platforms such as the FLU-IDIGM ACCESS ARRAY (48 reactions per sample in microfluidic chips) or DROPLET PCR by RAIN DANCE TECHNOLOGY (100 s to a few thousands of targets). Unfortunately, these split-and-pool methods are problematic for samples with a limited amount of DNA, as there is often not enough copies of the genome to ensure that there is one copy of each region of the genome in each well.

This is an especially severe problem when polymorphic loci are targeted, and the relative proportions of the alleles at the polymorphic loci are needed, as the stochastic noise introduced by the splitting and pooling will cause very poorly accurate measurements of the proportions of the alleles that were present in the original sample of DNA. Described here is a method to effectively and efficiently 40 amplify many PCR reactions that is applicable to cases where only a limited amount of DNA is available. In an embodiment, the method may be applied for analysis of single cells, body fluids, mixtures of DNA such as the free floating DNA found in maternal plasma, biopsies, environmental and/or forensic samples.

In an embodiment, the targeted sequencing may involve one, a plurality, or all of the following steps. a) Generate and amplify a library with adaptor sequences on both ends of DNA fragments. b) Divide into multiple reactions after 50 library amplification. c) Generate and optionally amplify a library with adaptor sequences on both ends of DNA fragments. d) Perform 1000- to 10,000-plex amplification of selected targets using one target specific "Forward" primer per target and one tag specific primer. e) Perform a second 55 amplification from this product using "Reverse" target specific primers and one (or more) primer specific to a universal tag that was introduced as part of the target specific forward primers in the first round. f) Perform a 1000-plex preamplification of selected target for a limited number of cycles. g) 60 Divide the product into multiple aliquots and amplify subpools of targets in individual reactions (for example, 50 to 500-plex, though this can be used all the way down to singleplex. h) Pool products of parallel subpools reactions. i) During these amplifications primers may carry sequencing 65 compatible tags (partial or full length) such that the products can be sequenced.

86

Highly Multiplexed PCR

Disclosed herein are methods that permit the targeted amplification of over a hundred to tens of thousands of target sequences (e.g., SNP loci) from a nucleic acid sample such as genomic DNA obtained from plasma. The amplified sample may be relatively free of primer dimer products and have low allelic bias at target loci. If during or after amplification the products are appended with sequencing compatible adaptors, analysis of these products can be performed by sequencing.

Performing a highly multiplexed PCR amplification using methods known in the art results in the generation of primer dimer products that are in excess of the desired amplification products and not suitable for sequencing. These can be reduced empirically by eliminating primers that form these products, or by performing in silico selection of primers. However, the larger the number of assays, the more difficult this problem becomes.

One solution is to split the 5000-plex reaction into several lower-plexed amplifications, e.g. one hundred 50-plex or 20 fifty 100-plex reactions, or to use microfluidics or even to split the sample into individual PCR reactions. However, if the sample DNA is limited, such as in non-invasive prenatal diagnostics from pregnancy plasma, dividing the sample between multiple reactions should be avoided as this will 25 result in bottlenecking.

Described herein are methods to first globally amplify the plasma DNA of a sample and then divide the sample up into multiple multiplexed target enrichment reactions with more moderate numbers of target sequences per reaction. In an embodiment, a method of the present disclosure can be used for preferentially enriching a DNA mixture at a plurality of loci, the method comprising one or more of the following steps: generating and amplifying a library from a mixture of DNA where the molecules in the library have adaptor sequences ligated on both ends of the DNA fragments, dividing the amplified library into multiple reactions, performing a first round of multiplex amplification of selected targets using one target specific "forward" primer per target and one or a plurality of adaptor specific universal "reverse' primers. In an embodiment, a method of the present disclosure further includes performing a second amplification using "reverse" target specific primers and one or a plurality of primers specific to a universal tag that was introduced as part of the target specific forward primers in the first round. In an embodiment, the method may involve a fully nested, hemi-nested, semi-nested, one sided fully nested, one sided hemi-nested, or one sided semi-nested PCR approach. In an embodiment, a method of the present disclosure is used for preferentially enriching a DNA mixture at a plurality of loci, the method comprising performing a multiplex preamplification of selected targets for a limited number of cycles, dividing the product into multiple aliquots and amplifying subpools of targets in individual reactions, and pooling products of parallel subpools reactions. Note that this approach could be used to perform targeted amplification in a manner that would result in low levels of allelic bias for 50-500 loci, for 500 to 5,000 loci, for 5,000 to 50,000 loci, or even for 50,000 to 500,000 loci. In an embodiment, the primers carry partial or full length sequencing compatible

The workflow may entail (1) extracting DNA such as plasma DNA, (2) preparing fragment library with universal adaptors on both ends of fragments, (3) amplifying the library using universal primers specific to the adaptors, (4) dividing the amplified sample "library" into multiple aliquots, (5) performing multiplex (e.g. about 100-plex, 1,000, or 10,000-plex with one target specific primer per target and

#### US 11,519,035 B2

8

a tag-specific primer) amplifications on aliquots, (6) pooling aliquots of one sample, (7) barcoding the sample, (8) mixing the samples and adjusting the concentration, (9) sequencing the sample. The workflow may comprise multiple sub-steps that contain one of the listed steps (e.g. step (2) of preparing the library step could entail three enzymatic steps (blunt ending, dA tailing and adaptor ligation) and three purification steps). Steps of the workflow may be combined, divided up or performed in different order (e.g. bar coding and pooling of samples).

It is important to note that the amplification of a library can be performed in such a way that it is biased to amplify short fragments more efficiently. In this manner it is possible to preferentially amplify shorter sequences, e.g. mononucleosomal DNA fragments as the cell free fetal DNA (of 15 placental origin) found in the circulation of pregnant women. Note that PCR assays can have the tags, for example sequencing tags, (usually a truncated form of 15-25 bases).

After multiplexing, PCR multiplexes of a sample are pooled and then the tags are completed (including bar 20 coding) by a tag-specific PCR (could also be done by ligation). Also, the full sequencing tags can be added in the same reaction as the multiplexing. In the first cycles targets may be amplified with the target specific primers, subsequently the tag-specific primers take over to complete the 25 SQ-adaptor sequence. The PCR primers may carry no tags. The sequencing tags may be appended to the amplification products by ligation.

In an embodiment, highly multiplex PCR followed by evaluation of amplified material by clonal sequencing may 30 be used for various applications such as the detection of fetal aneuploidy. Whereas traditional multiplex PCRs evaluate up to fifty loci simultaneously, the approach described herein may be used to enable simultaneous evaluation of more than 50 loci simultaneously, more than 100 loci simultaneously, more than 500 loci simultaneously, more than 1,000 loci simultaneously, more than 5,000 loci simultaneously, more than 10,000 loci simultaneously, more than 50,000 loci simultaneously, and more than 100,000 loci simultaneously. Experiments have shown that up to, including and more than 40 10,000 distinct loci can be evaluated simultaneously, in a single reaction, with sufficiently good efficiency and specificity to make non-invasive prenatal aneuploidy diagnoses and/or copy number calls with high accuracy. Assays may be combined in a single reaction with the entirety of a sample 4. such as a cfDNA sample isolated from maternal plasma, a fraction thereof, or a further processed derivative of the cfDNA sample. The sample (e.g., cfDNA or derivative) may also be split into multiple parallel multiplex reactions. The optimum sample splitting and multiplex is determined by 50 trading off various performance specifications. Due to the limited amount of material, splitting the sample into multiple fractions can introduce sampling noise, handling time, and increase the possibility of error. Conversely, higher multiplexing can result in greater amounts of spurious 55 amplification and greater inequalities in amplification both of which can reduce test performance.

Two crucial related considerations in the application of the methods described herein are the limited amount of original sample (e.g., plasma) and the number of original 60 molecules in that material from which allele frequency or other measurements are obtained. If the number of original molecules falls below a certain level, random sampling noise becomes significant, and can affect the accuracy of the test. Typically, data of sufficient quality for making noninvasive prenatal aneuploidy diagnoses can be obtained if measurements are made on a sample comprising the equiva-

QΩ

lent of 500-1000 original molecules per target locus. There are a number of ways of increasing the number of distinct measurements, for example increasing the sample volume. Each manipulation applied to the sample also potentially results in losses of material. It is essential to characterize losses incurred by various manipulations and avoid, or as necessary improve yield of certain manipulations to avoid losses that could degrade performance of the test.

In an embodiment, it is possible to mitigate potential losses in subsequent steps by amplifying all or a fraction of the original sample (e.g., cfDNA sample). Various methods are available to amplify all of the genetic material in a sample, increasing the amount available for downstream procedures. In an embodiment, ligation mediated PCR (LM-PCR) DNA fragments are amplified by PCR after ligation of either one distinct adaptors, two distinct adapters, or many distinct adaptors. In an embodiment, multiple displacement amplification (MDA) phi-29 polymerase is used to amplify all DNA isothermally. In DOP-PCR and variations, random priming is used to amplify the original material DNA. Each method has certain characteristics such as uniformity of amplification across all represented regions of the genome, efficiency of capture and amplification of original DNA, and amplification performance as a function of the length of the fragment.

In an embodiment LM-PCR may be used with a single heteroduplexed adaptor having a 3-prime tyrosine. The heteroduplexed adaptor enables the use of a single adaptor molecule that may be converted to two distinct sequences on 5-prime and 3-prime ends of the original DNA fragment during the first round of PCR. In an embodiment, it is possible to fractionate the amplified library by size separations, or products such as AMPURE, TASS or other similar methods. Prior to ligation, sample DNA may be blunt ended, and then a single adenosine base is added to the 3-prime end. Prior to ligation the DNA may be cleaved using a restriction enzyme or some other cleavage method. During ligation the 3-prime adenosine of the sample fragments and the complementary 3-prime tyrosine overhang of adaptor can enhance ligation efficiency. The extension step of the PCR amplification may be limited from a time standpoint to reduce amplification from fragments longer than about 200 bp, about 300 bp, about 400 bp, about 500 bp or about 1,000 bp. Since longer DNA found in the maternal plasma is nearly exclusively maternal, this may result in the enrichment of fetal DNA by 10-50% and improvement of test performance. A number of reactions were run using conditions as specified by commercially available kits; the resulted in successful ligation of fewer than 10% of sample DNA molecules. A series of optimizations of the reaction conditions for this improved ligation to approximately 70%. Mini-PCR

The following Mini-PCR method is desirable for samples containing short nucleic acids, digested nucleic acids, or fragmented nucleic acids, such as cfDNA. Traditional PCR assay design results in significant losses of distinct fetal molecules, but losses can be greatly reduced by designing very short PCR assays, termed mini-PCR assays. Fetal cfDNA in maternal serum is highly fragmented and the fragment sizes are distributed in approximately a Gaussian fashion with a mean of 160 bp, a standard deviation of 15 bp, a minimum size of about 100 bp, and a maximum size of about 220 bp. The distribution of fragment start and end positions with respect to the targeted polymorphisms, while not necessarily random, vary widely among individual targets and among all targets collectively and the polymorphic site of one particular target locus may occupy any position

#### US 11,519,035 B2

89

from the start to the end among the various fragments originating from that locus. Note that the term mini-PCR may equally well refer to normal PCR with no additional restrictions or limitations.

During PCR, amplification will only occur from template 5 DNA fragments comprising both forward and reverse primer sites. Because fetal cfDNA fragments are short, the likelihood of both primer sites being present the likelihood of a fetal fragment of length L comprising both the forward and reverse primers sites is ratio of the length of the amplicon to 10 the length of the fragment. Under ideal conditions, assays in which the amplicon is 45, 50, 55, 60, 65, or 70 bp will successfully amplify from 72%, 69%, 66%, 63%, 59%, or 56%, respectively, of available template fragment molecules. The amplicon length is the distance between the 15 5-prime ends of the forward and reverse priming sites. Amplicon length that is shorter than typically used by those known in the art may result in more efficient measurements of the desired polymorphic loci by only requiring short sequence reads. In an embodiment, a substantial fraction of 20 the amplicons should be less than 100 bp, less than 90 bp, less than 80 bp, less than 70 bp, less than 65 bp, less than 60 bp, less than 55 bp, less than 50 bp, or less than 45 bp.

Note that in methods known in the prior art, short assays such as those described herein are usually avoided because 25 they are not required and they impose considerable constraint on primer design by limiting primer length, annealing characteristics, and the distance between the forward and reverse primer.

Also note that there is the potential for biased amplifica- 30 tion if the 3-prime end of the either primer is within roughly 1-6 bases of the polymorphic site. This single base difference at the site of initial polymerase binding can result in preferential amplification of one allele, which can alter observed allele frequencies and degrade performance. All of 35 these constraints make it very challenging to identify primers that will amplify a particular locus successfully and furthermore, to design large sets of primers that are compatible in the same multiplex reaction. In an embodiment, the 3' end of the inner forward and reverse primers are 40 designed to hybridize to a region of DNA upstream from the polymorphic site, and separated from the polymorphic site by a small number of bases. Ideally, the number of bases may be between 6 and 10 bases, but may equally well be between 4 and 15 bases, between three and 20 bases, 45 between two and 30 bases, or between 1 and 60 bases, and achieve substantially the same end.

Multiplex PCR may involve a single round of PCR in which all targets are amplified or it may involve one round of PCR followed by one or more rounds of nested PCR or 50 some variant of nested PCR. Nested PCR consists of a subsequent round or rounds of PCR amplification using one or more new primers that bind internally, by at least one base pair, to the primers used in a previous round. Nested PCR reduces the number of spurious amplification targets by 55 amplifying, in subsequent reactions, only those amplification products from the previous one that have the correct internal sequence. Reducing spurious amplification targets improves the number of useful measurements that can be obtained, especially in sequencing. Nested PCR typically 60 entails designing primers completely internal to the previous primer binding sites, necessarily increasing the minimum DNA segment size required for amplification. For samples such as maternal plasma cfDNA, in which the DNA is highly fragmented, the larger assay size reduces the number of 65 distinct cfDNA molecules from which a measurement can be obtained. In an embodiment, to offset this effect, one may

90

use a partial nesting approach where one or both of the second round primers overlap the first binding sites extending internally some number of bases to achieve additional specificity while minimally increasing in the total assay size.

In an embodiment, a multiplex pool of PCR assays are designed to amplify potentially heterozygous SNP or other polymorphic or non-polymorphic loci on one or more chromosomes and these assays are used in a single reaction to amplify DNA. The number of PCR assays may be between 50 and 200 PCR assays, between 200 and 1,000 PCR assays, between 1,000 and 5,000 PCR assays, or between 5,000 and 20,000 PCR assays (50 to 200-plex, 200 to 1,000-plex, 1,000 to 5,000-plex, 5,000 to 20,000-plex, more than 20,000-plex respectively). In an embodiment, a multiplex pool of about 10,000 PCR assays (10,000-plex) are designed to amplify potentially heterozygous SNP loci on chromosomes X, Y, 13, 18, and 21 and 1 or 2 and these assays are used in a single reaction to amplify cfDNA obtained from a material plasma sample, chorion villus samples, amniocentesis samples, single or a small number of cells, other bodily fluids or tissues, cancers, or other genetic matter. The SNP frequencies of each locus may be determined by clonal or some other method of sequencing of the amplicons. Statistical analysis of the allele frequency distributions or ratios of all assays may be used to determine if the sample contains a trisomy of one or more of the chromosomes included in the test. In another embodiment the original cfDNA samples is split into two samples and parallel 5,000-plex assays are performed. In another embodiment the original cfDNA samples is split into n samples and parallel (~10,000/n)-plex assays are performed where n is between 2 and 12, or between 12 and 24, or between 24 and 48, or between 48 and 96. Data is collected and analyzed in a similar manner to that already described. Note that this method is equally well applicable to detecting translocations, deletions, duplications, and other chromosomal abnormalities.

In an embodiment, tails with no homology to the target genome may also be added to the 3-prime or 5-prime end of any of the primers. These tails facilitate subsequent manipulations, procedures, or measurements. In an embodiment, the tail sequence can be the same for the forward and reverse target specific primers. In an embodiment, different tails may be used for the forward and reverse target specific primers. In an embodiment, a plurality of different tails may be used for different loci or sets of loci. Certain tails may be shared among all loci or among subsets of loci. For example, using forward and reverse tails corresponding to forward and reverse sequences required by any of the current sequencing platforms can enable direct sequencing following amplification. In an embodiment, the tails can be used as common priming sites among all amplified targets that can be used to add other useful sequences. In some embodiments, the inner primers may contain a region that is designed to hybridize either upstream or downstream of the targeted locus (e.g, a polymorphic locus). In some embodiments, the primers may contain a molecular barcode. In some embodiments, the primer may contain a universal priming sequence designed to allow PCR amplification.

In an embodiment, a 10,000-plex PCR assay pool is created such that forward and reverse primers have tails corresponding to the required forward and reverse sequences required by a high throughput sequencing instrument such as the HISEQ, GAIIX, or MYSEQ available from ILLUMINA. In addition, included 5-prime to the sequencing tails is an additional sequence that can be used as a priming site in a subsequent PCR to add nucleotide barcode sequences to the

#### US 11,519,035 B2

9

amplicons, enabling multiplex sequencing of multiple samples in a single lane of the high throughput sequencing instrument.

In an embodiment, a 10,000-plex PCR assay pool is created such that reverse primers have tails corresponding to 5 the required reverse sequences required by a high throughput sequencing instrument. After amplification with the first 10,000-plex assay, a subsequent PCR amplification may be performed using a another 10,000-plex pool having partly nested forward primers (e.g. 6-bases nested) for all targets and a reverse primer corresponding to the reverse sequencing tail included in the first round. This subsequent round of partly nested amplification with just one target specific primer and a universal primer limits the required size of the assay, reducing sampling noise, but greatly reduces the 15 number of spurious amplicons. The sequencing tags can be added to appended ligation adaptors and/or as part of PCR probes, such that the tag is part of the final amplicon.

Fetal fraction affects performance of the test. There are a number of ways to enrich the fetal fraction of the DNA 20 found in maternal plasma. Fetal fraction can be increased by the previously described LM-PCR method already discussed as well as by a targeted removal of long maternal fragments. In an embodiment, prior to multiplex PCR amplification of the target loci, an additional multiplex PCR reaction may be 25 carried out to selectively remove long and largely maternal fragments corresponding to the loci targeted in the subsequent multiplex PCR. Additional primers are designed to anneal a site a greater distance from the polymorphism than is expected to be present among cell free fetal DNA frag- 30 ments. These primers may be used in a one cycle multiplex PCR reaction prior to multiplex PCR of the target polymorphic loci. These distal primers are tagged with a molecule or moiety that can allow selective recognition of the tagged pieces of DNA. In an embodiment, these molecules of DNA 3: may be covalently modified with a biotin molecule that allows removal of newly formed double stranded DNA comprising these primers after one cycle of PCR. Double stranded DNA formed during that first round is likely maternal in origin. Removal of the hybrid material may be 40 accomplish by the used of magnetic streptavidin beads. There are other methods of tagging that may work equally well. In an embodiment, size selection methods may be used to enrich the sample for shorter strands of DNA; for example those less than about 800 bp, less than about 500 bp, or less 45 than about 300 bp. Amplification of short fragments can then proceed as usual.

The mini-PCR method described in this disclosure enables highly multiplexed amplification and analysis of hundreds to thousands or even millions of loci in a single 5 reaction, from a single sample. At the same, the detection of the amplified DNA can be multiplexed; tens to hundreds of samples can be multiplexed in one sequencing lane by using barcoding PCR. This multiplexed detection has been successfully tested up to 49-plex, and a much higher degree of 55 multiplexing is possible. In effect, this allows hundreds of samples to be genotyped at thousands of SNPs in a single sequencing run. For these samples, the method allows determination of genotype and heterozygosity rate and simultaneously determination of copy number, both of 60 which may be used for the purpose of aneuploidy detection. This method is particularly useful in detecting aneuploidy of a gestating fetus from the free floating DNA found in maternal plasma. This method may be used as part of a method for sexing a fetus, and/or predicting the paternity of 65 the fetus. It may be used as part of a method for mutation dosage. This method may be used for any amount of DNA

92

or RNA, and the targeted regions may be SNPs, other polymorphic regions, non-polymorphic regions, and combinations thereof.

In some embodiments, ligation mediated universal-PCR amplification of fragmented DNA may be used. The ligation mediated universal-PCR amplification can be used to amplify plasma DNA, which can then be divided into multiple parallel reactions. It may also be used to preferentially amplify short fragments, thereby enriching fetal fraction. In some embodiments the addition of tags to the fragments by ligation can enable detection of shorter fragments, use of shorter target sequence specific portions of the primers and/or annealing at higher temperatures which reduces unspecific reactions.

The methods described herein may be used for a number of purposes where there is a target set of DNA that is mixed with an amount of contaminating DNA. In some embodiments, the target DNA and the contaminating DNA may be from individuals who are genetically related. For example, genetic abnormalities in a fetus (target) may be detected from maternal plasma which contains fetal (target) DNA and also maternal (contaminating) DNA; the abnormalities include whole chromosome abnormalities (e.g. aneuploidy) partial chromosome abnormalities (e.g. deletions, duplications, inversions, and translocations), polynucleotide polymorphisms (e.g. STRs), single nucleotide polymorphisms, and/or other genetic abnormalities or differences. In some embodiments, the target and contaminating DNA may be from the same individual, but where the target and contaminating DNA are different by one or more mutations, for example in the case of cancer. (see e.g. H. Mamon et al. Preferential Amplification of Apoptotic DNA from Plasma: Potential for Enhancing Detection of Minor DNA Alterations in Circulating DNA. Clinical Chemistry 54:9 (2008). In some embodiments, the DNA may be found in cell culture (apoptotic) supernatant. In some embodiments, it is possible to induce apoptosis in biological samples (e.g., blood) for subsequent library preparation, amplification and/or sequencing. A number of enabling workflows and protocols to achieve this end are presented elsewhere in this disclo-

In some embodiments, the target DNA may originate from single cells, from samples of DNA consisting of less than one copy of the target genome, from low amounts of DNA, from DNA from mixed origin (e.g. pregnancy plasma: placental and maternal DNA; cancer patient plasma and tumors: mix between healthy and cancer DNA, transplantation etc), from other body fluids, from cell cultures, from culture supernatants, from forensic samples of DNA, from ancient samples of DNA (e.g. insects trapped in amber), from other samples of DNA, and combinations thereof.

In some embodiments, a short amplicon size may be used. Short amplicon sizes are especially suited for fragmented DNA (see e.g. A. Sikora, et sl. Detection of increased amounts of cell-free fetal DNA with short PCR amplicons. Clin Chem. 2010 January; 56(1):136-8.)

The use of short amplicon sizes may result in some significant benefits. Short amplicon sizes may result in optimized amplification efficiency. Short amplicon sizes typically produce shorter products, therefore there is less chance for nonspecific priming. Shorter products can be clustered more densely on sequencing flow cell, as the clusters will be smaller. Note that the methods described herein may work equally well for longer PCR amplicons. Amplicon length may be increased if necessary, for example, when sequencing larger sequence stretches.

#### US 11,519,035 B2

9

Experiments with 146-plex targeted amplification with assays of 100 bp to 200 bp length as first step in a nested-PCR protocol were run on single cells and on genomic DNA with positive results.

In some embodiments, the methods described herein may 5 be used to amplify and/or detect SNPs, copy number, nucleotide methylation, mRNA levels, other types of RNA expression levels, other genetic and/or epigenetic features. The mini-PCR methods described herein may be used along with next-generation sequencing; it may be used with other 10 downstream methods such as microarrays, counting by digital PCR, real-time PCR, Mass-spectrometry analysis etc.

In some embodiment, the mini-PCR amplification methods described herein may be used as part of a method for accurate quantification of minority populations. It may be 15 used for absolute quantification using spike calibrators. It may be used for mutation/minor allele quantification through very deep sequencing, and may be run in a highly multiplexed fashion. It may be used for standard paternity and identity testing of relatives or ancestors, in human, 20 animals, plants or other creatures. It may be used for forensic testing. It may be used for rapid genotyping and copy number analysis (CN), on any kind of material, e.g. amniotic fluid and CVS, sperm, product of conception (POC). It may be used for single cell analysis, such as genotyping on 25 samples biopsied from embryos. It may be used for rapid embryo analysis (within less than one, one, or two days of biopsy) by targeted sequencing using min-PCR.

In some embodiments, it may be used for tumor analysis: tumor biopsies are often a mixture of health and tumor cells. 30 Targeted PCR allows deep sequencing of SNPs and loci with close to no background sequences. It may be used for copy number and loss of heterozygosity analysis on tumor DNA. Said tumor DNA may be present in many different body fluids or tissues of tumor patients. It may be used for 35 detection of tumor recurrence, and/or tumor screening. It may be used for quality control testing of seeds. It may be used for breeding, or fishing purposes. Note that any of these methods could equally well be used targeting non-polymorphic loci for the purpose of ploidy calling.

Some literature describing some of the fundamental methods that underlie the methods disclosed herein include: (1) Wang H Y, Luo M, Tereshchenko I V, Frikker D M, Cui X, Li J Y, Hu G, Chu Y, Azaro M A, Lin Y, Shen L, Yang Q, Kambouris M E, Gao R, Shih W, Li H. Genome Res. 2005 45 February; 15(2):276-83. Department of Molecular Genetics, Microbiology and Immunology/The Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, N.J. 08903, USA. (2) High-throughput genotyping of single nucleotide polymorphisms with high sensitivity. Li H, 50 Wang H Y, Cui X, Luo M, Hu G, Greenawalt D M, Tereshchenko I V, Li J Y, Chu Y, Gao R. Methods Mol Biol. 2007; 396—PubMed PMID: 18025699. (3) A method comprising multiplexing of an average of 9 assays for sequencing is described in: Nested Patch PCR enables highly 55 multiplexed mutation discovery in candidate genes. Varley K E, Mitra R D. Genome Res. 2008 November; 18(11): 1844-50. Epub 2008 Oct. 10. Note that the methods disclosed herein allow multiplexing of orders of magnitude more than in the above references.

Targeted PCR Variants—Nesting

There are many workflows that are possible when conducting PCR; some workflows typical to the methods disclosed herein are described. The steps outlined herein are not meant to exclude other possible steps nor does it imply that 65 any of the steps described herein are required for the method to work properly. A large number of parameter variations or

94

other modifications are known in the literature, and may be made without affecting the essence of the invention. One particular generalized workflow is given below followed by a number of possible variants. The variants typically refer to possible secondary PCR reactions, for example different types of nesting that may be done (step 3). It is important to note that variants may be done at different times, or in different orders than explicitly described herein. Examples that use polymorphic loci for illustration can be readily adapted for the amplification of nonpolymorphic loci if desired.

The DNA in the sample may have ligation adapters, often referred to as library tags or ligation adaptor tags (LTs), appended, where the ligation adapters contain a universal priming sequence, followed by a universal amplification. In an embodiment, this may be done using a standard protocol designed to create sequencing libraries after fragmentation. In an embodiment, the DNA sample can be blunt ended, and then an A can be added at the 3' end. A Y-adaptor with a T-overhang can be added and ligated. In some embodiments, other sticky ends can be used other than an A or T overhang. In some embodiments, other adaptors can be added, for example looped ligation adaptors. In some embodiments, the adaptors may have tag designed for PCR amplification.

Specific Target Amplification (STA): Pre-amplification of hundreds to thousands to tens of thousands and even hundreds of thousands of targets may be multiplexed in one reaction volume. STA is typically run from 10 to 30 cycles, though it may be run from 5 to 40 cycles, from 2 to 50 cycles, and even from 1 to 100 cycles. Primers may be tailed, for example for a simpler workflow or to avoid sequencing of a large proportion of dimers. Note that typically, dimers of both primers carrying the same tag will not be amplified or sequenced efficiently. In some embodiments, between 1 and 10 cycles of PCR may be carried out; in some embodiments between 10 and 20 cycles of PCR may be carried out; in some embodiments between 20 and 30 cycles of PCR may be carried out: in some embodiments between 30 and 40 cycles of PCR may be carried out; in some embodiments more than 40 cycles of PCR may be carried out. The amplification may be a linear amplification. The number of PCR cycles may be optimized to result in an optimal depth of read (DOR) profile. Different DOR profiles may be desirable for different purposes. In some embodiments, a more even distribution of reads between all assays is desirable; if the DOR is too small for some assays, the stochastic noise can be too high for the data to be too useful, while if the depth of read is too high, the marginal usefulness of each additional read is relatively small.

Primer tails may improve the detection of fragmented DNA from universally tagged libraries. If the library tag and the primer-tails contain a homologous sequence, hybridization can be improved (for example, melting temperature  $(T_M)$  is lowered) and primers can be extended if only a portion of the primer target sequence is in the sample DNA fragment. In some embodiments, 13 or more target specific base pairs may be used. In some embodiments, 10 to 12 target specific base pairs may be used. In some embodiments, 8 to 9 target specific base pairs may be used. In some embodiments, 6 to 7 target specific base pairs may be used. In some embodiments, STA may be performed on preamplified DNA, e.g. MDA, RCA, other whole genome amplifications, or adaptor-mediated universal PCR. In some embodiments, STA may be performed on samples that are enriched or depleted of certain sequences and populations, e.g. by size selection, target capture, directed degradation.

#### US 11,519,035 B2

9

In some embodiments, it is possible to perform secondary multiplex PCRs or primer extension reactions to increase specificity and reduce undesirable products. For example, full nesting, semi-nesting, hemi-nesting, and/or subdividing into parallel reactions of smaller assay pools are all techniques that may be used to increase specificity. Experiments have shown that splitting a sample into three 400-plex reactions resulted in product DNA with greater specificity than one 1,200-plex reaction with exactly the same primers. Similarly, experiments have shown that splitting a sample into four 2,400-plex reactions resulted in product DNA with greater specificity than one 9,600-plex reaction with exactly the same primers. In an embodiment, it is possible to use target-specific and tag specific primers of the same and opposing directionality.

In some embodiments, it is possible to amplify a DNA sample (dilution, purified or otherwise) produced by an STA reaction using tag-specific primers and "universal amplification", i.e. to amplify many or all pre-amplified and tagged targets. Primers may contain additional functional 20 sequences, e.g. barcodes, or a full adaptor sequence necessary for sequencing on a high throughput sequencing platform.

These methods may be used for analysis of any sample of DNA, and are especially useful when the sample of DNA is 25 particularly small, or when it is a sample of DNA where the DNA originates from more than one individual, such as in the case of maternal plasma. These methods may be used on DNA samples such as a single or small number of cells, genomic DNA, plasma DNA, amplified plasma libraries, 30 amplified apoptotic supernatant libraries, or other samples of mixed DNA. In an embodiment, these methods may be used in the case where cells of different genetic constitution may be present in a single individual, such as with cancer or transplants. In an embodiment, some of the DNA is from the recipient of a transplant (such as recipient cell-free or cellular DNA) and some of the DNA is from the donor of the transplant (such as cell-free or cellular DNA from the transplant). In an embodiment, the method is used to amplify one or more loci that differ between the recipient and the 40 donor (such as loci for which a different combination of alleles are present in the recipient compared to the donor). In some embodiments, the recipient is homozygous for a first allele (such as AA) and the donor is homozygous for a second allele (such as BB) or is heterozygous with the first 45 allele and a second allele (such as AB) at one or more loci. In some embodiments, the method is used to measure the absolute or relative amount of DNA from the donor of the transplant (such as cell-free or cellular DNA from the transplant). In some embodiments, this method is used to 50 prognose, diagnose, detect, or monitor a transplant status or outcome, such as transplant rejection, tolerance, non-rejection based allograft injury, transplant function, transplant survival, chronic transplant injury, or tittering of pharmacological immunosuppression.

Protocol Variants (Variants and/or Additions to the Workflow Above)

Direct multiplexed mini-PCR: Specific target amplification (STA) of a plurality of target sequences with tagged primers is shown in FIG. 1. 101 denotes double stranded 60 DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA that has been universally amplified with PCR primers hybridized. 104 denotes the final PCR product. In some 65 embodiments, STA may be done on more than 100, more than 200, more than 500, more than 1,000, more than 2,000,

96

more than 5,000, more than 10,000, more than 20,000, more than 50,000, more than 100,000 or more than 200,000 targets. In a subsequent reaction, tag-specific primers amplify all target sequences and lengthen the tags to include all necessary sequences for sequencing, including sample indexes. In an embodiment, primers may not be tagged or only certain primers may be tagged. Sequencing adaptors may be added by conventional adaptor ligation. In an embodiment, the initial primers may carry the tags.

In an embodiment, primers are designed so that the length of DNA amplified is unexpectedly short. Prior art demonstrates that ordinary people skilled in the art typically design 100+ bp amplicons. In an embodiment, the amplicons may be designed to be less than 80 bp. In an embodiment, the amplicons may be designed to be less than 70 bp. In an embodiment, the amplicons may be designed to be less than 60 bp. In an embodiment, the amplicons may be designed to be less than 50 bp. In an embodiment, the amplicons may be designed to be less than 45 bp. In an embodiment, the amplicons may be designed to be less than 40 bp. In an embodiment, the amplicons may be designed to be less than 35 bp. In an embodiment, the amplicons may be designed to be less than 35 bp. In an embodiment, the amplicons may be designed to be less than 45 bp. In an embodiment, the amplicons may be designed to be less than 45 bp. In an embodiment, the amplicons may be designed to be between 40 and 65 bp.

An experiment was performed using this protocol using 1200-plex amplification. Both genomic DNA and pregnancy plasma were used; about 70% of sequence reads mapped to targeted sequences. Details are given elsewhere in this document. Sequencing of a 1042-plex without design and selection of assays resulted in >99% of sequences being primer dimer products.

Sequential PCR:

After STA1 multiple aliquots of the product may be amplified in parallel with pools of reduced complexity with the same primers. The first amplification can give enough material to split. This method is especially good for small samples, for example those that are about 6-100 µg, about 100 µg to 1 ng, about 1 ng to 10 ng, or about 10 ng to 100 ng. The protocol was performed with 1200-plex into three 400-plexes. Mapping of sequencing reads increased from around 60 to 70% in the 1200-plex alone to over 95%.

Semi-Nested Mini-PCR:

(see FIG. 2) After STA 1 a second STA is performed comprising a multiplex set of internal nested Forward primers (103 B, 105 b) and one (or few) tag-specific Reverse primers (103 A). 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA that has been universally amplified with Forward primer B and Reverse Primer A hybridized. 104 denotes the PCR product from 103. 105 denotes the product from 104 with nested Forward primer b hybridized, and Reverse tag A already part of the molecule from the PCR that occurred between 103 and 104. 106 denotes the final PCR product. With this workflow usually greater than 95% of sequences map to the intended targets. The nested primer may overlap with the outer Forward primer sequence but introduces additional 3'-end bases. In some embodiments it is possible to use between one and 20 extra 3' bases. Experiments have shown that using 9 or more extra 3' bases in a 1200-plex designs works well. As readily apparent, the primers for the second STA can alternatively be considered a multiplex set of internal nested Reverse primers and one (or a few) tagspecific Forward primers

Fully Nested Mini-PCR:

(see FIG. 3) After STA step 1, it is possible to perform a second multiplex PCR (or parallel m.p. PCRs of reduced

#### US 11,519,035 B2

9'

complexity) with two nested primers carrying tags (A, a, B, b). 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA that has been universally 5 amplified with Forward primer B and Reverse Primer A hybridized. 104 denotes the PCR product from 103. 105 denotes the product from 104 with nested Forward primer b and nested Reverse primer a hybridized. 106 denotes the final PCR product. In some embodiments, it is possible to use two full sets of primers. Experiments using a fully nested mini-PCR protocol were used to perform 146-plex amplification on single and three cells without step 102 of appending universal ligation adaptors and amplifying.

Hemi-Nested Mini-PCR:

(see FIG. 4) It is possible to use target DNA that has and adaptors at the fragment ends. STA is performed comprising a multiplex set of Forward primers (B) and one (or few) tag-specific Reverse primers (A). A second STA can be performed using a universal tag-specific Forward primer and 20 target specific Reverse primer. 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA that has been universally amplified with Reverse 25 Primer A hybridized. 104 denotes the PCR product from 103 that was amplified using Reverse primer A and ligation adaptor tag primer LT. 105 denotes the product from 104 with Forward primer B hybridized. 106 denotes the final PCR product. In this workflow, target specific Forward and 30 Reverse primers are used in separate reactions, thereby reducing the complexity of the reaction and preventing dimer formation of forward and reverse primers. Note that in this example, primers A and B may be considered to be first primers, and primers 'a' and 'b' may be considered to 3 be inner primers. This method is a big improvement on direct PCR as it is as good as direct PCR, but it avoids primer dimers. After first round of hemi nested protocol one typically sees 99% non-targeted DNA, however, after second round there is typically a big improvement. As readily 40 apparent, the primers for the first STA can be considered a multiplex set of Reverse primers and one (or few) tagspecific Forward primers, and the primers for the second STA can be considered a universal tag-specific Reverse primer and target specific Forward primer(s).

Triply Hemi-Nested Mini-PCR:

(see FIG. 5) It is possible to use target DNA that has and adaptor at the fragment ends. STA is performed comprising a multiplex set of Forward primers (B) and one (or few) tag-specific Reverse primers (A) and (a). A second STA can 50 be performed using a universal tag-specific Forward primer and target specific Reverse primer. 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single 55 stranded DNA that has been universally amplified with Reverse Primer A hybridized. 104 denotes the PCR product from 103 that was amplified using Reverse primer A and ligation adaptor tag primer LT. 105 denotes the product from 104 with Forward primer B hybridized. 106 denotes the PCR 60 product from 105 that was amplified using Reverse primer A and Forward primer B. 107 denotes the product from 106 with Reverse primer 'a' hybridized. 108 denotes the final PCR product. Note that in this example, primers 'a' and B may be considered to be inner primers, and A may be 65 considered to be a first primer. Optionally, both A and B may be considered to be first primers, and 'a' may be considered

98

to be an inner primer. The designation of reverse and forward primers may be switched. In this workflow, target specific Forward and Reverse primers are used in separate reactions, thereby reducing the complexity of the reaction and preventing dimer formation of forward and reverse primers. This method is a big improvement on direct PCR as it is as good as direct PCR, but it avoids primer dimers. After first round of hemi nested protocol one typically sees 99% non-targeted DNA, however, after second round there is typically a big improvement.

One-Sided Nested Mini-PCR:

(see FIG. 6) It is possible to use target DNA that has an adaptor at the fragment ends. STA may also be performed with a multiplex set of nested Forward primers and using the 15 ligation adapter tag as the Reverse primer. A second STA may then be performed using a set of nested Forward primers and a universal Reverse primer. 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA that has been universally amplified with Forward Primer A hybridized. 104 denotes the PCR product from 103 that was amplified using Forward primer A and ligation adaptor tag Reverse primer LT. 105 denotes the product from 104 with nested Forward primer a hybridized. 106 denotes the final PCR product. This method can detect shorter target sequences than standard PCR by using overlapping primers in the first and second STAs. The method is typically performed off a sample of DNA that has already undergone STA step 1 above-appending of universal tags and amplification; the two nested primers are only on one side, other side uses the library tag. The method was performed on libraries of apoptotic supernatants and pregnancy plasma. With this workflow around 60% of sequences mapped to the intended targets. Note that reads that contained the reverse adaptor sequence were not mapped, so this number is expected to be higher if those reads that contain the reverse adaptor sequence are mapped

One-Sided Mini-PCR:

It is possible to use target DNA that has an adaptor at the fragment ends (see FIG. 7). STA may be performed with a multiplex set of Forward primers and one (or few) tagspecific Reverse primer. 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA with Forward Primer A hybridized. 104 denotes the PCR product from 103 that was amplified using Forward primer A and ligation adaptor tag Reverse primer LT, and which is the final PCR product. This method can detect shorter target sequences than standard PCR. However it may be relatively unspecific, as only one target specific primer is used. This protocol is effectively half of the one sided nested mini PCR

Reverse Semi-Nested Mini-PCR:

It is possible to use target DNA that has an adaptor at the fragment ends (see FIG. 8). STA may be performed with a multiplex set of Forward primers and one (or few) tagspecific Reverse primer. 101 denotes double stranded DNA with a polymorphic locus of interest at X. 102 denotes the double stranded DNA with ligation adaptors added for universal amplification. 103 denotes the single stranded DNA with Reverse Primer B hybridized. 104 denotes the PCR product from 103 that was amplified using Reverse primer B and ligation adaptor tag Forward primer LT. 105 denotes the PCR product 104 with hybridized Forward Primer A, and inner Reverse primer 'b'. 106 denotes the

Case: 24-1324 Document: 42-1 Page: 172 Filed: 03/18/2024

#### US 11,519,035 B2

99

PCR product that has been amplified from 105 using Forward primer A and Reverse primer 'b', and which is the final PCR product. This method can detect shorter target sequences than standard PCR.

There also may be more variants that are simply iterations or combinations of the above methods such as doubly nested PCR, where three sets of primers are used. Another variant is one-and-a-half sided nested mini-PCR, where STA may also be performed with a multiplex set of nested Forward primers and one (or few) tag-specific Reverse primer.

Note that in all of these variants, the identity of the Forward primer and the Reverse primer may be interchanged. Note that in some embodiments, the nested variant can equally well be run without the initial library preparation  $_{15}$ that comprises appending the adapter tags, and a universal amplification step. Note that in some embodiments, additional rounds of PCR may be included, with additional Forward and/or Reverse primers and amplification steps; these additional steps may be particularly useful if it is 20 desirable to further increase the percent of DNA molecules that correspond to the targeted loci.

There are many ways to perform the amplification, with different degrees of nesting, and with different degrees of 25 multiplexing. In FIG. 9, a flow chart is given with some of the possible workflows. Note that the use of 10,000-plex PCR is only meant to be an example; these flow charts would work equally well for other degrees of multiplexing. Looped Ligation Adaptors

Nesting Workflows

When adding universal tagged adaptors for example for the purpose of making a library for sequencing, there are a number of ways to ligate adaptors. One way is to blunt end the sample DNA, perform A-tailing, and ligate with adaptors that have a T-overhang. There are a number of other ways to ligate adaptors. There are also a number of adaptors that can be ligated. For example, a Y-adaptor can be used where the adaptor consists of two strands of DNA where one strand has a double strand region, and a region specified by a forward primer region, and where the other strand specified by a 40 double strand region that is complementary to the double strand region on the first strand, and a region with a reverse primer. The double stranded region, when annealed, may contain a T-overhang for the purpose of ligating to double stranded DNA with an A overhang.

In an embodiment, the adaptor can be a loop of DNA where the terminal regions are complementary, and where the loop region contains a forward primer tagged region (LFT), a reverse primer tagged region (LRT), and a cleavage site between the two (See FIG. 10). 101 refers to the double 50 stranded, blunt ended target DNA. 102 refers to the A-tailed target DNA. 103 refers to the looped ligation adaptor with T overhang 'T' and the cleavage site 'Z'. 104 refers to the target DNA with appended looped ligation adaptors. 105 refers to the target DNA with the ligation adaptors appended 55 cleaved at the cleavage site. LFT refers to the ligation adaptor Forward tag, and the LRT refers to the ligation adaptor Reverse tag. The complementary region may end on a T overhang, or other feature that may be used for ligation to the target DNA. The cleavage site may be a series of 60 uracils for cleavage by UNG, or a sequence that may be recognized and cleaved by a restriction enzyme or other method of cleavage or just a basic amplification. These adaptors can be uses for any library preparation, for example, for sequencing. These adaptors can be used in 65 combination with any of the other methods described herein, for example the mini-PCR amplification methods.

100

Internally Tagged Primers

When using sequencing to determine the allele present at given polymorphic locus, the sequence read typically begins upstream of the primer binding site (a), and then to the polymorphic site (X). Tags are typically configured as shown in FIG. 11, left. 101 refers to the single stranded target DNA with polymorphic locus of interest 'X', and primer 'a' with appended tag 'b'. In order to avoid nonspecific hybridization, the primer binding site (region of target DNA complementary to 'a') is typically 18 to 30 bp in length. Sequence tag 'b' is typically about 20 bp; in theory these can be any length longer than about 15 bp, though many people use the primer sequences that are sold by the sequencing platform company. The distance 'd' between 'a' and 'X' may be at least 2 bp so as to avoid allele bias. When performing multiplexed PCR amplification using the methods disclosed herein or other methods, where careful primer design is necessary to avoid excessive primer primer interaction, the window of allowable distance 'd' between 'a' and X' may vary quite a bit: from 2 bp to 10 bp, from 2 bp to 20 bp, from 2 bp to 30 bp, or even from 2 bp to more than 30 bp. Therefore, when using the primer configuration shown in FIG. 11, left, sequence reads must be a minimum of 40 bp to obtain reads long enough to measure the polymorphic locus, and depending on the lengths of 'a' and 'd' the sequence reads may need to be up to 60 or 75 bp. Usually, the longer the sequence reads, the higher the cost and time of sequencing a given number of reads, therefore, minimizing the necessary read length can save both time and money. In addition, since, on average, bases read earlier on the read are read more accurately than those read later on the read, decreasing the necessary sequence read length can also increase the accuracy of the measurements of the polymorphic region.

In an embodiment, termed internally tagged primers, the primer binding site (a) is split in to a plurality of segments (a', a", a'"  $\dots$  ), and the sequence tag (b) is on a segment of DNA that is in the middle of two of the primer binding sites, as shown in FIG. 11, 103. This configuration allows the sequencer to make shorter sequence reads. In an embodiment, a'+a" should be at least about 18 bp, and can be as long as 30, 40, 50, 60, 80, 100 or more than 100 bp. In an embodiment, a" should be at least about 6 bp, and in an embodiment is between about 8 and 16 bp. All other factors being equal, using the internally tagged primers can cut the length of the sequence reads needed by at least 6 bp, as much as 8 bp, 10 bp, 12 bp, 15 bp, and even by as many as 20 or 30 bp. This can result in a significant money, time and accuracy advantage. An example of internally tagged primers is given in FIG. 12.

Primers with Ligation Adaptor Binding Region

One issue with fragmented DNA is that since it is short in length, the chance that a polymorphism is close to the end of a DNA strand is higher than for a long strand (e.g. 101, FIG. 10). Since PCR capture of a polymorphism requires a primer binding site of suitable length on both sides of the polymorphism, a significant number of strands of DNA with the targeted polymorphism will be missed due to insufficient overlap between the primer and the targeted binding site. In an embodiment, the target DNA 101 can have ligation adaptors appended 102, and the target primer 103 can have a region (cr) that is complementary to the ligation adaptor tag (lt) appended upstream of the designed binding region (a) (see FIG. 13); thus in cases where the binding region (region of 101 that is complementary to a) is shorter than the 18 bp typically required for hybridization, the region (cr) on the primer than is complementary to the library tag is able to increase the binding energy to a point where the PCR can

#### US 11,519,035 B2

10

proceed. Note that any specificity that is lost due to a shorter binding region can be made up for by other PCR primers with suitably long target binding regions. Note that this embodiment can be used in combination with direct PCR, or any of the other methods described herein, such as nested PCR, semi nested PCR, hemi nested PCR, one sided nested or semi or hemi nested PCR, or other PCR protocols.

When using the sequencing data to determine ploidy in combination with an analytical method that involves comparing the observed allele data to the expected allele distri- 10 butions for various hypotheses, each additional read from alleles with a low depth of read will yield more information than a read from an allele with a high depth of read. Therefore, ideally, one would wish to see uniform depth of read (DOR) where each locus will have a similar number of 15 representative sequence reads. Therefore, it is desirable to minimize the DOR variance. In an embodiment, it is possible to decrease the coefficient of variance of the DOR (this may be defined as the standard deviation of the DOR/the average DOR) by increasing the annealing times. In some 20 embodiments the annealing temperatures may be longer than 2 minutes, longer than 4 minutes, longer than ten minutes, longer than 30 minutes, and longer than one hour, or even longer. Since annealing is an equilibrium process, there is no limit to the improvement of DOR variance with increasing 25 and/or displaced. annealing times. In an embodiment, increasing the primer concentration may decrease the DOR variance **Exemplary Amplification Methods** 

Improved PCR amplification methods have also been developed that minimize or prevent interference due to the 30 amplification of nearby or adjacent target loci in the same reaction volume (such as part of the sample multiplex PCR reaction that simultaneously amplifies all the target loci) (see, U.S. Ser. No. 61/982,245, filed Apr. 21, 2014; U.S. Ser. No. 61/987,407, filed May 1, 2014, and U.S. Ser. No. 3 62/066,514, filed Oct. 21, 2014, which are each hereby incorporated by reference in its entirety). These methods can be used to simultaneously amplify nearby or adjacent target loci, which is faster and cheaper than having to separate nearby target loci into different reaction volumes so that they 40 can be amplified separately to avoid interference. In particular embodiments, these methods are used to tile a region such that the amplicons include all the nucleotides in the region (such as an exon or all the exons of a gene such as cystic fibrosis).

In some embodiments, the amplification of target loci is performed using a polymerase (e.g., a DNA polymerase, RNA polymerase, or reverse transcriptase) with low 5'→3' exonuclease and/or low strand displacement activity. In some embodiments, the low level of 5'→3' exonuclease 50 reduces or prevents the degradation of a nearby primer (e.g., an unextended primer or a primer that has had one or more nucleotides added to during primer extension). In some embodiments, the low level of strand displacement activity reduces or prevents the displacement of a nearby primer 55 (e.g., an unextended primer or a primer that has had one or more nucleotides added to it during primer extension). In some embodiments, target loci that are adjacent to each other (e.g., no bases between the target loci) or nearby (e.g., loci are within 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, 60 or 1 base) are amplified. In some embodiments, the 3' end of one locus is within 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 base of the 5' end of next downstream locus.

In some embodiments, at least 100, 200, 500, 750, 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 65 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified, such as by the

102

simultaneous amplification in one reaction volume In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products are target amplicons. In various embodiments, the amount of amplified products that are target amplicons is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 98%, 90 to 99.5%, or 95 to 99.5%, inclusive. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the target loci are amplified (e.g., amplified at least 5, 10, 20, 30, 50, or 100-fold compared to the amount prior to amplification), such as by the simultaneous amplification in one reaction volume. In various embodiments, the amount target loci that are amplified (e.g, amplified at least 5, 10, 20, 30, 50, or 100-fold compared to the amount prior to amplification) is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 99%, 90 to 99.5%, 95 to 99.9%, or 98 to 99.99% inclusive. In some embodiments, fewer non-target amplicons are produced, such as fewer amplicons formed from a forward primer from a first primer pair and a reverse primer from a second primer pair. Such undesired non-target amplicons can be produced using prior amplification methods if, e.g., the reverse primer from the first primer pair and/or the forward primer from the second primer pair are degraded

In some embodiments, these methods allows longer extension times to be used since the polymerase bound to a primer being extended is less likely to degrade and/or displace a nearby primer (such as the next downstream primer) given the low 5'→3' exonuclease and/or low strand displacement activity of the polymerase. In various embodiments, reaction conditions (such as the extension time and temperature) are used such that the extension rate of the polymerase allows the number of nucleotides that are added to a primer being extended to be equal to or greater than 80, 90, 95, 100, 110, 120, 130, 140, 150, 175, or 200% of the number of nucleotides between the 3' end of the primer binding site and the 5'end of the next downstream primer binding site on the same strand.

In some embodiments, a DNA polymerase is used produce DNA amplicons using DNA as a template. In some embodiments, a RNA polymerase is used produce RNA amplicons using DNA as a template. In some embodiments, a reverse transcriptase is used produce cDNA amplicons using RNA as a template.

In some embodiments, the low level of 5'→3' exonuclease of the polymerase is less than 80, 70, 60, 50, 40, 30, 20, 10, 5, 1, or 0.1% of the activity of the same amount of *Thermus aquaticus* polymerase ("Taq" polymerase, which is a commonly used DNA polymerase from a thermophilic bacterium, PDB 1BGX, EC 2.7.7.7, Murali et al., "Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme," Proc. Natl. Acad. Sci. USA 95:12562-12567, 1998, which is hereby incorporated by reference in its entirety) under the same conditions. In some embodiments, the low level of strand displacement activity of the polymerase is less than 80, 70, 60, 50, 40, 30, 20, 10, 5, 1, or 0.1% of the activity of the same amount of Taq polymerase under the same conditions.

In some embodiments, the polymerase is a PUSHION DNA polymerase, such as PHUSION High Fidelity DNA polymerase (M0530S, New England BioLabs, Inc.) or PHUSION Hot Start Flex DNA polymerase (M0535S, New England BioLabs, Inc.; Frey and Suppman BioChemica. 2:34-35, 1995; Chester and Marshak *Analytical Biochemistry*. 209:284-290, 1993, which are each hereby incorporated

#### US 11,519,035 B2

10.

by reference in its entirety). The PHUSION DNA polymerase is a *Pyrococcus*-like enzyme fused with a processivity-enhancing domain.

PHUSION DNA polymerase possesses 5'→3' polymerase activity and 3'→5' exonuclease activity, and generates bluntended products. PHUSION DNA polymerase lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, the polymerase is a Q5® DNA Polymerase, such as Q5® High-Fidelity DNA Polymerase (M0491S, New England BioLabs, Inc.) or Q5® Hot Start High-Fidelity DNA Polymerase (M0493S, New England BioLabs, Inc.). Q5® High-Fidelity DNA polymerase is a high-fidelity, thermostable, DNA polymerase with 3'→5' exonuclease activity, fused to a processivity-enhancing Sso7d domain. Q5® High-Fidelity DNA polymerase lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, the polymerase is a T4 DNA polymerase (M0203S, New England BioLabs, Inc.; Tabor and Struh. (1989). "DNA-Dependent DNA Polymerases," In 20 Ausebel et al. (Ed.), Current Protocols in Molecular Biology. 3.5.10-3.5.12. New York: John Wiley & Sons, Inc., 1989; Sambrook et al. Molecular Cloning: A Laboratory Manual. (2nd ed.), 5.44-5.47. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989, which are each 25 hereby incorporated by reference in its entirety). T4 DNA Polymerase catalyzes the synthesis of DNA in the 5'→3' direction and requires the presence of template and primer. This enzyme has a 3'→5' exonuclease activity which is much more active than that found in DNA Polymerase I. T4 30 DNA polymerase lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, the polymerase is a *Sulfolobus* DNA Polymerase IV (M0327S, New England BioLabs, Inc.; (Boudsocq, et al. (2001). *Nucleic Acids Res.*, 29:4607-4616, 35 2001; McDonald, et al. (2006). *Nucleic Acids Res.*, 34:1102-1111, 2006, which are each hereby incorporated by reference in its entirety). *Sulfolobus* DNA Polymerase IV is a thermostable Y-family lesion-bypass DNA Polymerase that efficiently synthesizes DNA across a variety of DNA template 40 lesions McDonald, J. P. et al. (2006). *Nucleic Acids Res.*, 34, 1102-1111, which is hereby incorporated by reference in its entirety). *Sulfolobus* DNA Polymerase IV lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, if a primer binds a region with a 45 SNP, the primer may bind and amplify the different alleles with different efficiencies or may only bind and amplify one allele. For subjects who are heterozygous, one of the alleles may not be amplified by the primer. In some embodiments, a primer is designed for each allele. For example, if there are 50 two alleles (e.g., a biallelic SNP), then two primers can be used to bind the same location of a target locus (e.g., a forward primer to bind the "A" allele and a forward primer to bind the "B" allele). Standard methods, such as the dbSNP database, can be used to determine the location of known 55 SNPs, such as SNP hot spots that have a high heterozygosity rate.

In some embodiments, the amplicons are similar in size. In some embodiments, the range of the length of the target amplicons is less than 100, 75, 50, 25, 15, 10, or 5 nucleotides. In some embodiments (such as the amplification of target loci in fragmented DNA or RNA), the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 and 75 nucleotides, inclusive. In some embodiments (such as the amplification 65 of multiple target loci throughout an exon or gene), the length of the target amplicons is between 100 and 500

104

nucleotides, such as between 150 and 450 nucleotides, 200 and 400 nucleotides, 200 and 300 nucleotides, or 300 and 400 nucleotides, inclusive.

In some embodiments, multiple target loci are simultaneously amplified using a primer pair that includes a forward and reverse primer for each target locus to be amplified in that reaction volume. In some embodiments, one round of PCR is performed with a single primer per target locus, and then a second round of PCR is performed with a primer pair per target locus. For example, the first round of PCR may be performed with a single primer per target locus such that all the primers bind the same strand (such as using a forward primer for each target locus). This allows the PCR to amplify in a linear manner and reduces or eliminates amplification bias between amplicons due to sequence or length differences. In some embodiments, the amplicons are then amplified using a forward and reverse primer for each target locus. Exemplary Whole Genome Amplification Methods

In some embodiments, a method of the present disclosure may involve amplifying DNA, such as the use of whole genome application to amplify a nucleic acid sample before amplifying just the target loci. Amplification of the DNA, a process which transforms a small amount of genetic material to a larger amount of genetic material that comprises a similar set of genetic data, can be done by a wide variety of methods, including, but not limited to polymerase chain reaction (PCR). One method of amplifying DNA is whole genome amplification (WGA). There are a number of methods available for WGA: ligation-mediated PCR (LM-PCR), degenerate oligonucleotide primer PCR (DOP-PCR), and multiple displacement amplification (MDA). In LM-PCR, short DNA sequences called adapters are ligated to blunt ends of DNA. These adapters contain universal amplification sequences, which are used to amplify the DNA by PCR. In DOP-PCR, random primers that also contain universal amplification sequences are used in a first round of annealing and PCR. Then, a second round of PCR is used to amplify the sequences further with the universal primer sequences. MDA uses the phi-29 polymerase, which is a highly processive and non-specific enzyme that replicates DNA and has been used for single-cell analysis. The major limitations to amplification of material from a single cell are (1) necessity of using extremely dilute DNA concentrations or extremely small volume of reaction mixture, and (2) difficulty of reliably dissociating DNA from proteins across the whole genome. Regardless, single-cell whole genome amplification has been used successfully for a variety of applications for a number of years. There are other methods of amplifying DNA from a sample of DNA. The DNA amplification transforms the initial sample of DNA into a sample of DNA that is similar in the set of sequences, but of much greater quantity. In some cases, amplification may not be required.

In some embodiments, DNA may be amplified using a universal amplification, such as WGA or MDA. In some embodiments, DNA may be amplified by targeted amplification, for example using targeted PCR, or circularizing probes. In some embodiments, the DNA may be preferentially enriched using a targeted amplification method, or a method that results in the full or partial separation of desired from undesired DNA, such as capture by hybridization approaches. In some embodiments, DNA may be amplified by using a combination of a universal amplification method and a preferential enrichment method. A fuller description of some of these methods can be found elsewhere in this document.

#### US 11,519,035 B2

105

Exemplary Enrichment and Sequencing Methods

In an embodiment, a method disclosed herein uses selective enrichment techniques that preserve the relative allele frequencies that are present in the original sample of DNA at each target loci (e.g., each polymorphic locus) from a set 5 of target loci (e.g., polymorphic loci). While enrichment is particularly advantageous for methods for analyzing polymorphic loci, these enrichment methods can be readily adapted for nonpolymorphic loci if desired. In some embodiments the amplification and/or selective enrichment 10 technique may involve PCR such as ligation mediated PCR, fragment capture by hybridization, Molecular Inversion Probes, or other circularizing probes. In some embodiments, methods for amplification or selective enrichment may involve using probes where, upon correct hybridization to 15 the target sequence, the 3-prime end or 5-prime end of a nucleotide probe is separated from the polymorphic site of the allele by a small number of nucleotides. This separation reduces preferential amplification of one allele, termed allele bias. This is an improvement over methods that involve 20 using probes where the 3-prime end or 5-prime end of a correctly hybridized probe are directly adjacent to or very near to the polymorphic site of an allele. In an embodiment, probes in which the hybridizing region may or certainly contains a polymorphic site are excluded. Polymorphic sites 25 at the site of hybridization can cause unequal hybridization or inhibit hybridization altogether in some alleles, resulting in preferential amplification of certain alleles. These embodiments are improvements over other methods that involve targeted amplification and/or selective enrichment 30 in that they better preserve the original allele frequencies of the sample at each polymorphic locus, whether the sample is pure genomic sample from a single individual or mixture

The use of a technique to enrich a sample of DNA at a set 35 of target loci followed by sequencing as part of a method for non-invasive prenatal allele calling or ploidy calling may confer a number of unexpected advantages. In some embodiments of the present disclosure, the method involves measuring genetic data for use with an informatics based 40 method, such as PARENTAL SUPPORT<sup>TM</sup> (PS). The ultimate outcome of some of the embodiments is the actionable genetic data of an embryo or a fetus. There are many methods that may be used to measure the genetic data of the individual and/or the related individuals as part of embodied 45 methods. In an embodiment, a method for enriching the concentration of a set of targeted alleles is disclosed herein, the method comprising one or more of the following steps: targeted amplification of genetic material, addition of loci specific oligonucleotide probes, ligation of specified DNA 50 strands, isolation of sets of desired DNA, removal of unwanted components of a reaction, detection of certain sequences of DNA by hybridization, and detection of the sequence of one or a plurality of strands of DNA by DNA sequencing methods. In some cases the DNA strands may 55 refer to target genetic material, in some cases they may refer to primers, in some cases they may refer to synthesized sequences, or combinations thereof. These steps may be carried out in a number of different orders.

For example, a universal amplification step of the DNA 60 prior to targeted amplification may confer several advantages, such as removing the risk of bottlenecking and reducing allelic bias. The DNA may be mixed an oligonucleotide probe that can hybridize with two neighboring regions of the target sequence, one on either side. After 65 hybridization, the ends of the probe may be connected by adding a polymerase, a means for ligation, and any neces-

106

sary reagents to allow the circularization of the probe. After circularization, an exonuclease may be added to digest to non-circularized genetic material, followed by detection of the circularized probe. The DNA may be mixed with PCR primers that can hybridize with two neighboring regions of the target sequence, one on either side. After hybridization, the ends of the probe may be connected by adding a polymerase, a means for ligation, and any necessary reagents to complete PCR amplification. Amplified or unamplified DNA may be targeted by hybrid capture probes that target a set of loci; after hybridization, the probe may be localized and separated from the mixture to provide a mixture of DNA that is enriched in target sequences.

The use of a method to target certain loci followed by sequencing as part of a method for allele calling or ploidy calling may confer a number of unexpected advantages. Some methods by which DNA may be targeted, or preferentially enriched, include using circularizing probes, linked inverted probes (LIPs, MIPs), capture by hybridization methods such as SURESELECT, and targeted PCR or ligation-mediated PCR amplification strategies.

In some embodiments, a method of the present disclosure involves measuring genetic data for use with an informatics based method, such as PARENTAL SUPPORT<sup>TM</sup> (PS), which is described further herein. PARENTAL SUPPORT<sup>TM</sup> is an informatics based approach to manipulating genetic data, aspects of which are described herein. The ultimate outcome of some of the embodiments is the actionable genetic data of an embryo or a fetus followed by a clinical decision based on the actionable data. The algorithms behind the PS method take the measured genetic data of the target individual, often an embryo or fetus, and the measured genetic data from related individuals, and are able to increase the accuracy with which the genetic state of the target individual is known. In an embodiment, the measured genetic data is used in the context of making ploidy determinations during prenatal genetic diagnosis. In an embodiment, the measured genetic data is used in the context of making ploidy determinations or allele calls on embryos during in vitro fertilization. There are many methods that may be used to measure the genetic data of the individual and/or the related individuals in the aforementioned contexts. The different methods comprise a number of steps, those steps often involving amplification of genetic material, addition of oligonucleotide probes, ligation of specified DNA strands, isolation of sets of desired DNA, removal of unwanted components of a reaction, detection of certain sequences of DNA by hybridization, detection of the sequence of one or a plurality of strands of DNA by DNA sequencing methods. In some cases the DNA strands may refer to target genetic material, in some cases they may refer to primers, in some cases they may refer to synthesized sequences, or combinations thereof. These steps may be carried out in a number of different orders.

Note that in theory it is possible to target any number loci in the genome, anywhere from one loci to well over one million loci. If a sample of DNA is subjected to targeting, and then sequenced, the percentage of the alleles that are read by the sequencer will be enriched with respect to their natural abundance in the sample. The degree of enrichment can be anywhere from one percent (or even less) to ten-fold, a hundred-fold, a thousand-fold or even many million-fold. In the human genome there are roughly 3 billion base pairs, and nucleotides, comprising approximately 75 million polymorphic loci. The more loci that are targeted, the smaller the degree of enrichment is possible. The fewer the number of loci that are targeted, the greater degree of enrichment is

#### US 11,519,035 B2

107

possible, and the greater depth of read may be achieved at those loci for a given number of sequence reads.

In an embodiment of the present disclosure, the targeting or preferential may focus entirely on SNPs. In an embodiment, the targeting or preferential may focus on any polymorphic site. A number of commercial targeting products are available to enrich exons. Surprisingly, targeting exclusively SNPs, or exclusively polymorphic loci, is particularly advantageous when using a method for NPD that relies on allele distributions. There are also published methods for 10 NPD using sequencing, for example U.S. Pat. No. 7,888, 017, involving a read count analysis where the read counting focuses on counting the number of reads that map to a given chromosome, where the analyzed sequence reads do not focused on regions of the genome that are polymorphic. 15 Those types of methodology that do not focus on polymorphic alleles would not benefit as much from targeting or preferential enrichment of a set of alleles

In an embodiment of the present disclosure, it is possible to use a targeting method that focuses on SNPs to enrich a 20 genetic sample in polymorphic regions of the genome. In an embodiment, it is possible to focus on a small number of SNPs, for example between 1 and 100 SNPs, or a larger number, for example, between 100 and 1,000, between 1,000 and 10,000, between 10,000 and 100,000 or more than 25 100,000 SNPs. In an embodiment, it is possible to focus on one or a small number of chromosomes that are correlated with live trisomic births, for example chromosomes 13, 18, 21, X and Y, or some combination thereof. In an embodiment, it is possible to enrich the targeted SNPs by a small 30 factor, for example between 1.01 fold and 100 fold, or by a larger factor, for example between 100 fold and 1,000,000 fold, or even by more than 1,000,000 fold. In an embodiment of the present disclosure, it is possible to use a targeting method to create a sample of DNA that is preferentially enriched in polymorphic regions of the genome. In an embodiment, it is possible to use this method to create a mixture of DNA with any of these characteristics where the mixture of DNA contains maternal DNA and also free floating fetal DNA. In an embodiment, it is possible to use 40 this method to create a mixture of DNA that has any combination of these factors. For example, the method described herein may be used to produce a mixture of DNA that comprises maternal DNA and fetal DNA, and that is preferentially enriched in DNA that corresponds to 200 SNPs, all of which are located on either chromosome 18 or 21, and which are enriched an average of 1000 fold. In another example, it is possible to use the method to create a mixture of DNA that is preferentially enriched in 10,000 SNPs that are all or mostly located on chromosomes 13, 18, 50 21, X and Y, and the average enrichment per loci is greater than 500 fold. Any of the targeting methods described herein can be used to create mixtures of DNA that are preferentially enriched in certain loci.

In some embodiments, a method of the present disclosure 55 further includes measuring the DNA in the mixed fraction using a high throughput DNA sequencer, where the DNA in the mixed fraction contains a disproportionate number of sequences from one or more chromosomes, wherein the one or more chromosomes are taken from the group comprising 60 chromosome 13, chromosome 18, chromosome 21, chromosome X, chromosome Y and combinations thereof.

Described herein are three methods: multiplex PCR, targeted capture by hybridization, and linked inverted probes (LIPs), which may be used to obtain and analyze 65 measurements from a sufficient number of polymorphic loci from a maternal plasma sample in order to detect fetal

108

aneuploidy; this is not meant to exclude other methods of selective enrichment of targeted loci. Other methods may equally well be used without changing the essence of the method. In each case the polymorphism assayed may include single nucleotide polymorphisms (SNPs), small indels, or STRs. A preferred method involves the use of SNPs. Each approach produces allele frequency data; allele frequency data for each targeted locus and/or the joint allele frequency distributions from these loci may be analyzed to determine the ploidy of the fetus. Each approach has its own considerations due to the limited source material and the fact that maternal plasma consists of mixture of maternal and fetal DNA. This method may be combined with other approaches to provide a more accurate determination. In an embodiment, this method may be combined with a sequence counting approach such as that described in U.S. Pat. No. 7,888,017. The approaches described could also be used to detect fetal paternity noninvasively from maternal plasma samples. In addition each approach may be applied to other mixtures of DNA or pure DNA samples to detect the presence or absence of an euploid chromosomes, to genotype a large number of SNP from degraded DNA samples, to detect segmental copy number variations (CNVs), to detect other genotypic states of interest, or some combination thereof.

Accurately Measuring the Allelic Distributions in a Sample Current sequencing approaches can be used to estimate the distribution of alleles in a sample. One such method involves randomly sampling sequences from a pool DNA, termed shotgun sequencing. The proportion of a particular allele in the sequencing data is typically very low and can be determined by simple statistics. The human genome contains approximately 3 billion base pairs. So, if the sequencing method used make 100 bp reads, a particular allele will be measured about once in every 30 million sequence reads.

In an embodiment, a method of the present disclosure is used to determine the presence or absence of two or more different haplotypes that contain the same set of loci in a sample of DNA from the measured allele distributions of loci from that chromosome. The different haplotypes could represent two different homologous chromosomes from one individual, three different homologous chromosomes from a trisomic individual, three different homologous haplotypes from a mother and a fetus where one of the haplotypes is shared between the mother and the fetus, three or four haplotypes from a mother and fetus where one or two of the haplotypes are shared between the mother and the fetus, or other combinations. Alleles that are polymorphic between the haplotypes tend to be more informative, however any alleles where the mother and father are not both homozygous for the same allele will yield useful information through measured allele distributions beyond the information that is available from simple read count analysis.

Shotgun sequencing of such a sample, however, is extremely inefficient as it results in many sequences for regions that are not polymorphic between the different haplotypes in the sample, or are for chromosomes that are not of interest, and therefore reveal no information about the proportion of the target haplotypes. Described herein are methods that specifically target and/or preferentially enrich segments of DNA in the sample that are more likely to be polymorphic in the genome to increase the yield of allelic information obtained by sequencing. Note that for the measured allele distributions in an enriched sample to be truly representative of the actual amounts present in the target individual, it is critical that there is little or no preferential enrichment of one allele as compared to the other allele at a

#### US 11,519,035 B2

109

given loci in the targeted segments. Current methods known in the art to target polymorphic alleles are designed to ensure that at least some of any alleles present are detected. However, these methods were not designed for the purpose of measuring the unbiased allelic distributions of polymorphic alleles present in the original mixture. It is non-obvious that any particular method of target enrichment would be able to produce an enriched sample wherein the measured allele distributions would accurately represent the allele distributions present in the original unamplified sample better than any other method. While many enrichment methods may be expected, in theory, to accomplish such an aim, an ordinary person skilled in the art is well aware that there is a great deal of stochastic or deterministic bias in current amplification, targeting and other preferential 15 enrichment methods. One embodiment of a method described herein allows a plurality of alleles found in a mixture of DNA that correspond to a given locus in the genome to be amplified, or preferentially enriched in a way that the degree of enrichment of each of the alleles is nearly 20 the same. Another way to say this is that the method allows the relative quantity of the alleles present in the mixture as a whole to be increased, while the ratio between the alleles that correspond to each locus remains essentially the same as they were in the original mixture of DNA. For some reported 25 methods, preferential enrichment of loci can result in allelic biases of more than 1%, more than 2%, more than 5% and even more than 10%. This preferential enrichment may be due to capture bias when using a capture by hybridization approach, or amplification bias which may be small for each 30 cycle, but can become large when compounded over 20, 30 or 40 cycles. For the purposes of this disclosure, for the ratio to remain essentially the same means that the ratio of the alleles in the original mixture divided by the ratio of the alleles in the resulting mixture is between 0.95 and 1.05, 35 between 0.98 and 1.02, between 0.99 and 1.01, between 0.995 and 1.005, between 0.998 and 1.002, between 0.999 and 1.001, or between 0.9999 and 1.0001. Note that the calculation of the allele ratios presented here may not be used in the determination of the ploidy state of the target 40 individual, and may only a metric to be used to measure allelic bias

In an embodiment, once a mixture has been preferentially enriched at the set of target loci, it may be sequenced using any one of the previous, current, or next generation of 4. sequencing instruments that sequences a clonal sample (a sample generated from a single molecule; examples include ILLUMINA GAIIX, ILLUMINA HISEQ, LIFE TECH-NOLOGIES SOLiD, 5500XL). The ratios can be evaluated by sequencing through the specific alleles within the tar- 50 geted region. These sequencing reads can be analyzed and counted according the allele type and the rations of different alleles determined accordingly. For variations that are one to a few bases in length, detection of the alleles will be performed by sequencing and it is essential that the sequenc- 55 ing read span the allele in question in order to evaluate the allelic composition of that captured molecule. The total number of captured molecules assayed for the genotype can be increased by increasing the length of the sequencing read. Full sequencing of all molecules would guarantee collection 60 of the maximum amount of data available in the enriched pool. However, sequencing is currently expensive, and a method that can measure allele distributions using a lower number of sequence reads will have great value. In addition, there are technical limitations to the maximum possible 65 length of read as well as accuracy limitations as read lengths increase. The alleles of greatest utility will be of one to a few

110

bases in length, but theoretically any allele shorter than the length of the sequencing read can be used. While allele variations come in all types, the examples provided herein focus on SNPs or variants contained of just a few neighboring base pairs. Larger variants such as segmental copy number variants can be detected by aggregations of these smaller variations in many cases as whole collections of SNP internal to the segment are duplicated. Variants larger than a few bases, such as STRs require special consideration and some targeting approaches work while others will not.

There are multiple targeting approaches that can be used to specifically isolate and enrich a one or a plurality of variant positions in the genome. Typically, these rely on taking advantage of the invariant sequence flanking the variant sequence. There are reports by others related to targeting in the context of sequencing where the substrate is maternal plasma (see, e.g., Liao et al., Clin. Chem. 2011; 57(1): pp. 92-101). However, these approaches use targeting probes that target exons, and do not focus on targeting polymorphic regions of the genome. In an embodiment, a method of the present disclosure involves using targeting probes that focus exclusively or almost exclusively on polymorphic regions. In an embodiment, a method of the present disclosure involves using targeting probes that focus exclusively or almost exclusively on SNPs. In some embodiments of the present disclosure, the targeted polymorphic sites consist of at least 10% SNPs, at least 20% SNPs, at least 30% SNPs, at least 40% SNPs, at least 50% SNPs, at least 60% SNPs, at least 70% SNPs, at least 80% SNPs, at least 90% SNPs, at least 95% SNPs, at least 98% SNPs, at least 99% SNPs, at least 99.9% SNPs, or exclusively SNPs.

In an embodiment, a method of the present disclosure can be used to determine genotypes (base composition of the DNA at specific loci) and relative proportions of those genotypes from a mixture of DNA molecules, where those DNA molecules may have originated from one or a number of genetically distinct individuals. In an embodiment, a method of the present disclosure can be used to determine the genotypes at a set of polymorphic loci, and the relative ratios of the amount of different alleles present at those loci. In an embodiment the polymorphic loci may consist entirely of SNPs. In an embodiment, the polymorphic loci can comprise SNPs, single tandem repeats, and other polymorphisms. In an embodiment, a method of the present disclosure can be used to determine the relative distributions of alleles at a set of polymorphic loci in a mixture of DNA, where the mixture of DNA comprises DNA that originates from a mother, and DNA that originates from a fetus. In an embodiment, the joint allele distributions can be determined on a mixture of DNA isolated from blood from a pregnant woman. In an embodiment, the allele distributions at a set of loci can be used to determine the ploidy state of one or more chromosomes on a gestating fetus.

In an embodiment, the mixture of DNA molecules could be derived from DNA extracted from multiple cells of one individual. In an embodiment, the original collection of cells from which the DNA is derived may comprise a mixture of diploid or haploid cells of the same or of different genotypes, if that individual is mosaic (germline or somatic). In an embodiment, the mixture of DNA molecules could also be derived from DNA extracted from single cells. In an embodiment, the mixture of DNA molecules could also be derived from DNA extracted from mixture of two or more cells of the same individual, or of different individuals. In an embodiment, the mixture of DNA molecules could be derived from DNA isolated from biological material that has already liberated from cells such as blood plasma, which is

#### US 11,519,035 B2

111

known to contain cell free DNA. In an embodiment, the this biological material may be a mixture of DNA from one or more individuals, as is the case during pregnancy where it has been shown that fetal DNA is present in the mixture. In an embodiment, the biological material could be from a mixture of cells that were found in maternal blood, where some of the cells are fetal in origin. In an embodiment, the biological material could be cells from the blood of a pregnant which have been enriched in fetal cells. Circularizing Probes

Some embodiments of the present disclosure involve the use of "Linked Inverted Probes" (LIPs), which have been previously described in the literature, to amplify the target loci before or after amplification using primers that are not LIPs in the multiplex PCR methods of the invention. LIPs is a generic term meant to encompass technologies that involve the creation of a circular molecule of DNA, where the probes are designed to hybridize to targeted region of DNA on either side of a targeted allele, such that addition of 20 appropriate polymerases and/or ligases, and the appropriate conditions, buffers and other reagents, will complete the complementary, inverted region of DNA across the targeted allele to create a circular loop of DNA that captures the information found in the targeted allele. LIPs may also be 25 called pre-circularized probes, pre-circularizing probes, or circularizing probes. The LIPs probe may be a linear DNA molecule between 50 and 500 nucleotides in length, and in an embodiment between 70 and 100 nucleotides in length; in some embodiments, it may be longer or shorter than 30 described herein. Others embodiments of the present disclosure involve different incarnations, of the LIPs technology, such as Padlock Probes and Molecular Inversion Probes

One method to target specific locations for sequencing is 35 to synthesize probes in which the 3' and 5' ends of the probes anneal to target DNA at locations adjacent to and on either side of the targeted region, in an inverted manner, such that the addition of DNA polymerase and DNA ligase results in extension from the 3' end, adding bases to single stranded 40 probe that are complementary to the target molecule (gapfill), followed by ligation of the new 3' end to the 5' end of the original probe resulting in a circular DNA molecule that can be subsequently isolated from background DNA. The probe ends are designed to flank the targeted region of 45 interest. One aspect of this approach is commonly called MIPS and has been used in conjunction with array technologies to determine the nature of the sequence filled in. One drawback to the use of MIPs in the context of measuring allele ratios is that the hybridization, circularization and 50 amplification steps do not happed at equal rates for different alleles at the same loci. This results in measured allele ratios that are not representative of the actual allele ratios present in the original mixture.

In an embodiment, the circularizing probes are constructed such that the region of the probe that is designed to hybridize upstream of the targeted polymorphic locus and the region of the probe that is designed to hybridize downstream of the targeted polymorphic locus are covalently connected through a non-nucleic acid backbone. This backbone can be any biocompatible molecule or combination of biocompatible molecules. Some examples of possible biocompatible molecules are poly(ethylene glycol), polycarbonates, polyurethanes, polyethylenes, polypropylenes, sulfone polymers, silicone, cellulose, fluoropolymers, 65 acrylic compounds, styrene block copolymers, and other block copolymers.

112

In an embodiment of the present disclosure, this approach has been modified to be easily amenable to sequencing as a means of interrogating the filled in sequence. In order to retain the original allelic proportions of the original sample at least one key consideration must be taken into account. The variable positions among different alleles in the gap-fill region must not be too close to the probe binding sites as there can be initiation bias by the DNA polymerase resulting in differential of the variants. Another consideration is that additional variations may be present in the probe binding sites that are correlated to the variants in the gap-fill region which can result unequal amplification from different alleles. In an embodiment of the present disclosure, the 3' ends and 5' ends of the pre-circularized probe are designed to hybridize to bases that are one or a few positions away from the variant positions (polymorphic sites) of the targeted allele. The number of bases between the polymorphic site (SNP or otherwise) and the base to which the 3' end and/or 5' of the pre-circularized probe is designed to hybridize may be one base, it may be two bases, it may be three bases, it may be four bases, it may be five bases, it may be six bases, it may be seven to ten bases, it may be eleven to fifteen bases, or it may be sixteen to twenty bases, twenty to thirty bases, or thirty to sixty bases. The forward and reverse primers may be designed to hybridize a different number of bases away from the polymorphic site. Circularizing probes can be generated in large numbers with current DNA synthesis technology allowing very large numbers of probes to be generated and potentially pooled, enabling interrogation of many loci simultaneously. It has been reported to work with more than 300,000 probes. Two papers that discuss a method involving circularizing probes that can be used to measure the genomic data of the target individual include: Porreca et al., Nature Methods, 2007 4(11), pp. 931-936.; and also Turner et al., Nature Methods, 2009, 6(5), pp. 315-316. The methods described in these papers may be used in combination with other methods described herein. Certain steps of the method from these two papers may be used in combination with other steps from other methods described herein.

In some embodiments of the methods disclosed herein, the genetic material of the target individual is optionally amplified, followed by hybridization of the pre-circularized probes, performing a gap fill to fill in the bases between the two ends of the hybridized probes, ligating the two ends to form a circularized probe, and amplifying the circularized probe, using, for example, rolling circle amplification. Once the desired target allelic genetic information is captured by circularizing appropriately designed oligonucleotide probes, such as in the LIPs system, the genetic sequence of the circularized probes may be being measured to give the desired sequence data. In an embodiment, the appropriately designed oligonucleotides probes may be circularized directly on unamplified genetic material of the target individual, and amplified afterwards. Note that a number of amplification procedures may be used to amplify the original genetic material, or the circularized LIPs, including rolling circle amplification, MDA, or other amplification protocols. Different methods may be used to measure the genetic information on the target genome, for example using high throughput sequencing, Sanger sequencing, other sequencing methods, capture-by-hybridization, capture-by-circularization, multiplex PCR, other hybridization methods, and combinations thereof.

Once the genetic material of the individual has been measured using one or a combination of the above methods, an informatics based method, such as the PARENTAL

#### US 11,519,035 B2

113

SUPPORT<sup>TM</sup> method, along with the appropriate genetic measurements, can then be used to determination the ploidy state of one or more chromosomes on the individual, and/or the genetic state of one or a set of alleles, specifically those alleles that are correlated with a disease or genetic state of 5 interest. Note that the use of LIPs has been reported for multiplexed capture of genetic sequences, followed by genotyping with sequencing. However, the use of sequencing data resulting from a LIPs-based strategy for the amplification of the genetic material found in a single cell, a small 10 number of cells, or extracellular DNA, has not been used for the purpose of determining the ploidy state of a target individual.

Applying an informatics based method to determine the ploidy state of an individual from genetic data as measured 15 by hybridization arrays, such as the ILLUMINA INFINIUM array, or the AFFYMETRIX gene chip has been described in documents references elsewhere in this document. However, the method described herein shows improvements over methods described previously in the literature. For example, 20 the LIPs based approach followed by high throughput sequencing unexpectedly provides better genotypic data due to the approach having better capacity for multiplexing, better capture specificity, better uniformity, and low allelic bias. Greater multiplexing allows more alleles to be targeted, 25 giving more accurate results. Better uniformity results in more of the targeted alleles being measured, giving more accurate results. Lower rates of allelic bias result in lower rates of miscalls, giving more accurate results. More accurate results result in an improvement in clinical outcomes, 30 and better medical care.

It is important to note that LIPs may be used as a method for targeting specific loci in a sample of DNA for genotyping by methods other than sequencing. For example, LIPs may be used to target DNA for genotyping using SNP arrays or 35 other DNA or RNA based microarrays.

Ligation-Mediated PCR

Ligation-mediated PCR may be used to amplify the target loci before or after PCR amplification using primers that are not ligated. Ligation-mediated PCR is a method of PCR used 40 to preferentially enrich a sample of DNA by amplifying one or a plurality of loci in a mixture of DNA, the method comprising: obtaining a set of primer pairs, where each primer in the pair contains a target specific sequence and a non-target sequence, where the target specific sequence is 45 preferably designed to anneal to a target region, one upstream and one downstream from the polymorphic site, and which can be separated from the polymorphic site by 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-20, 21-30, 31-40, 41-50, 51-100, or more than 100; polymerization of the DNA from 50 the 3-prime end of upstream primer to the fill the single strand region between it and the 5-prime end of the downstream primer with nucleotides complementary to the target molecule; ligation of the last polymerized base of the upstream primer to the adjacent 5-prime base of the down- 55 stream primer; and amplification of only polymerized and ligated molecules using the non-target sequences contained at the 5-prime end of the upstream primer and the 3-prime end of the downstream primer. Pairs of primers to distinct targets may be mixed in the same reaction. The non-target 60 sequences serve as universal sequences such that of all pairs of primers that have been successfully polymerized and ligated may be amplified with a single pair of amplification

Capture by Hybridization

In some embodiments, a method of the present disclosure may involve using any of the following capture by hybrid-

114

ization methods in addition to using multiplex PCR to amplify the target loci. Preferential enrichment of a specific set of sequences in a target genome can be accomplished in a number of ways. Elsewhere in this document is a description of how LIPs can be used to target a specific set of sequences, but in all of those applications, other targeting and/or preferential enrichment methods can be used equally well for the same ends. One example of another targeting method is the capture by hybridization approach. Some examples of commercial capture by hybridization technologies include AGILENT's SURE SELECT and ILLUMI-NA's TRUSEQ. In capture by hybridization, a set of oligonucleotides that is complimentary or mostly complimentary to the desired targeted sequences is allowed to hybridize to a mixture of DNA, and then physically separated from the mixture. Once the desired sequences have hybridized to the targeting oligonucleotides, the effect of physically removing the targeting oligonucleotides is to also remove the targeted sequences. Once the hybridized oligos are removed, they can be heated to above their melting temperature and they can be amplified. Some ways to physically remove the targeting oligonucleotides is by covalently bonding the targeting oligos to a solid support, for example a magnetic bead, or a chip. Another way to physically remove the targeting oligonucleotides is by covalently bonding them to a molecular moiety with a strong affinity for another molecular moiety. An example of such a molecular pair is biotin and streptavidin, such as is used in SURE SELECT. Thus that targeted sequences could be covalently attached to a biotin molecule, and after hybridization, a solid support with streptavidin affixed can be used to pull down the biotinylated oligonucleotides, to which are hybridized to the targeted sequences

Hybrid capture involves hybridizing probes that are complementary to the targets of interest to the target molecules. Hybrid capture probes were originally developed to target and enrich large fractions of the genome with relative uniformity between targets. In that application, it was important that all targets be amplified with enough uniformity that all regions could be detected by sequencing, however, no regard was paid to retaining the proportion of alleles in original sample. Following capture, the alleles present in the sample can be determined by direct sequencing of the captured molecules. These sequencing reads can be analyzed and counted according the allele type. However, using the current technology, the measured allele distributions the captured sequences are typically not representative of the original allele distributions.

In an embodiment, detection of the alleles is performed by sequencing. In order to capture the allele identity at the polymorphic site, it is essential that the sequencing read span the allele in question in order to evaluate the allelic composition of that captured molecule. Since the capture molecules are often of variable lengths upon sequencing cannot be guaranteed to overlap the variant positions unless the entire molecule is sequenced. However, cost considerations as well as technical limitations as to the maximum possible length and accuracy of sequencing reads make sequencing the entire molecule unfeasible. In an embodiment, the read length can be increased from about 30 to about 50 or about 70 bases can greatly increase the number of reads that overlap the variant positions within the targeted sequences.

Another way to increase the number of reads that interrogate the position of interest is to decrease the length of the probe, as long as it does not result in bias in the underlying enriched alleles. The length of the synthesized probe should be long enough such that two probes designed to hybridize

#### US 11,519,035 B2

115

to two different alleles found at one locus will hybridize with near equal affinity to the various alleles in the original sample. Currently, methods known in the art describe probes that are typically longer than 120 bases. In a current embodiment, if the allele is one or a few bases then the capture probes may be less than about 110 bases, less than about 100 bases, less than about 90 bases, less than about 80 bases, less than about 70 bases, less than about 60 bases, less than about 50 bases, less than about 40 bases, less than about 30 bases, and less than about 25 bases, and this is sufficient to ensure 10 equal enrichment from all alleles. When the mixture of DNA that is to be enriched using the hybrid capture technology is a mixture comprising free floating DNA isolated from blood, for example maternal blood, the average length of DNA is quite short, typically less than 200 bases. The use of shorter 15 probes results in a greater chance that the hybrid capture probes will capture desired DNA fragments. Larger variations may require longer probes. In an embodiment, the variations of interest are one (a SNP) to a few bases in length. In an embodiment, targeted regions in the genome 20 can be preferentially enriched using hybrid capture probes wherein the hybrid capture probes are of a length below 90 bases, and can be less than 80 bases, less than 70 bases, less than 60 bases, less than 50 bases, less than 40 bases, less than 30 bases, or less than 25 bases. In an embodiment, to 25 increase the chance that the desired allele is sequenced, the length of the probe that is designed to hybridize to the regions flanking the polymorphic allele location can be decreased from above 90 bases, to about 80 bases, or to about 70 bases, or to about 60 bases, or to about 50 bases, 30 or to about 40 bases, or to about 30 bases, or to about 25 bases.

There is a minimum overlap between the synthesized probe and the target molecule in order to enable capture. This synthesized probe can be made as short as possible 35 while still being larger than this minimum required overlap. The effect of using a shorter probe length to target a polymorphic region is that there will be more molecules that overlap the target allele region. The state of fragmentation of the original DNA molecules also affects the number of reads 4 that will overlap the targeted alleles. Some DNA samples such as plasma samples are already fragmented due to biological processes that take place in vivo. However, samples with longer fragments by benefit from fragmentation prior to sequencing library preparation and enrichment. 45 When both probes and fragments are short (~60-80 bp) maximum specificity may be achieved relatively few sequence reads failing to overlap the critical region of interest

In an embodiment, the hybridization conditions can be 50 adjusted to maximize uniformity in the capture of different alleles present in the original sample. In an embodiment, hybridization temperatures are decreased to minimize differences in hybridization bias between alleles. Methods known in the art avoid using lower temperatures for hybrid- 55 ization because lowering the temperature has the effect of increasing hybridization of probes to unintended targets. However, when the goal is to preserve allele ratios with maximum fidelity, the approach of using lower hybridization temperatures provides optimally accurate allele ratios, 60 despite the fact that the current art teaches away from this approach. Hybridization temperature can also be increased to require greater overlap between the target and the synthesized probe so that only targets with substantial overlap of the targeted region are captured. In some embodiments of 65 the present disclosure, the hybridization temperature is lowered from the normal hybridization temperature to about 40°

116

C., to about  $45^{\circ}$  C., to about  $50^{\circ}$  C., to about  $55^{\circ}$  C., to about  $60^{\circ}$  C., to about  $60^{\circ}$  C.

In an embodiment, the hybrid capture probes can be designed such that the region of the capture probe with DNA that is complementary to the DNA found in regions flanking the polymorphic allele is not immediately adjacent to the polymorphic site. Instead, the capture probe can be designed such that the region of the capture probe that is designed to hybridize to the DNA flanking the polymorphic site of the target is separated from the portion of the capture probe that will be in van der Waals contact with the polymorphic site by a small distance that is equivalent in length to one or a small number of bases. In an embodiment, the hybrid capture probe is designed to hybridize to a region that is flanking the polymorphic allele but does not cross it; this may be termed a flanking capture probe. The length of the flanking capture probe may be less than about 120 bases, less than about 110 bases, less than about 100 bases, less than about 90 bases, and can be less than about 80 bases, less than about 70 bases, less than about 60 bases, less than about 50 bases, less than about 40 bases, less than about 30 bases, or less than about 25 bases. The region of the genome that is targeted by the flanking capture probe may be separated by the polymorphic locus by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-20, or more than 20 base pairs.

Description of a targeted capture based disease screening test using targeted sequence capture. Custom targeted sequence capture, like those currently offered by AGILENT (SURE SELECT), ROCHE-NIMBLEGEN, or ILLUMINA. Capture probes could be custom designed to ensure capture of various types of mutations. For point mutations, one or more probes that overlap the point mutation should be sufficient to capture and sequence the mutation.

For small insertions or deletions, one or more probes that overlap the mutation may be sufficient to capture and sequence fragments comprising the mutation. Hybridization may be less efficient between the probe-limiting capture efficiency, typically designed to the reference genome sequence. To ensure capture of fragments comprising the mutation one could design two probes, one matching the normal allele and one matching the mutant allele. A longer probe may enhance hybridization. Multiple overlapping probes may enhance capture. Finally, placing a probe immediately adjacent to, but not overlapping, the mutation may permit relatively similar capture efficiency of the normal and mutant alleles.

For Simple Tandem Repeats (STRs), a probe overlapping these highly variable sites is unlikely to capture the fragment well. To enhance capture a probe could be placed adjacent to, but not overlapping the variable site. The fragment could then be sequenced as normal to reveal the length and composition of the STR.

For large deletions, a series of overlapping probes, a common approach currently used in exon capture systems may work. However, with this approach it may be difficult to determine whether or not an individual is heterozygous. Targeting and evaluating SNPs within the captured region could potentially reveal loss of heterozygosity across the region indicating that an individual is a carrier. In an embodiment, it is possible to place non-overlapping or singleton probes across the potentially deleted region and use the number of fragments captured as a measure of heterozygosity. In the case where an individual caries a large deletion, one-half the number of fragments are expected to be available for capture relative to a non-deleted (diploid) reference locus. Consequently, the number of reads obtained from the deleted regions should be roughly half that

## US 11,519,035 B2

117

obtained from a normal diploid locus. Aggregating and averaging the sequencing read depth from multiple singleton probes across the potentially deleted region may enhance the signal and improve confidence of the diagnosis. The two approaches, targeting SNPs to identify loss of heterozygosity and using multiple singleton probes to obtain a quantitative measure of the quantity of underlying fragments from that locus can also be combined. Either or both of these strategies may be combined with other strategies to better obtain the same end.

If during testing cfDNA detection of a male fetus, as indicated by the presence of the Y-chromosome fragments, captured and sequenced in the same test, and either an X-linked dominant mutation where mother and father are unaffected, or a dominant mutation where mother is not 15 affected would indicated heighted risk to the fetus. Detection of two mutant recessive alleles within the same gene in an unaffected mother would imply the fetus had inherited a mutant allele from father and potentially a second mutant allele from mother. In all cases, follow-up testing by amniocentesis or chorionic villus sampling may be indicated.

A targeted capture based disease screening test could be combined with a targeted capture based non-invasive prenatal diagnostic test for an euploidy.

There are a number of ways to decrease depth of read 25 (DOR) variability: for example, one could increase primer concentrations, one could use longer targeted amplification probes, or one could run more STA cycles (such as more than 25, more than 30, more than 35, or even more than 40) Exemplary Methods of Determining the Number of DNA 30 Molecules in a Sample.

A method is described herein to determine the number of DNA molecules in a sample by generating a uniquely identified molecule for each original DNA molecules in the sample during the first round of DNA amplification. 3. Described here is a procedure to accomplish the above end followed by a single molecule or clonal sequencing method.

The approach entails targeting one or more specific loci and generating a tagged copy of the original molecules such manner that most or all of the tagged molecules from each 40 targeted locus will have a unique tag and can be distinguished from one another upon sequencing of this barcode using clonal or single molecule sequencing. Each unique sequenced barcode represents a unique molecule in the original sample. Simultaneously, sequencing data is used to 45 ascertain the locus from which the molecule originates. Using this information one can determine the number of unique molecules in the original sample for each locus.

This method can be used for any application in which quantitative evaluation of the number of molecules in an 50 original sample is required. Furthermore, the number of unique molecules of one or more targets can be related to the number of unique molecules to one or more other targets to determine the relative copy number, allele distribution, or allele ratio. Alternatively, the number of copies detected 55 from various targets can be modeled by a distribution in order to identify the mostly likely number of copies of the original targets. Applications include but are not limited to detection of insertions and deletions such as those found in carriers of Duchenne Muscular Dystrophy; quantitation of 60 deletions or duplications segments of chromosomes such as those observed in copy number variants; chromosome copy number of samples from born individuals; chromosome copy number of samples from unborn individuals such as embryos or fetuses

The method can be combined with simultaneous evaluation of variations contained in the targeted by sequence. This 118

can be used to determine the number of molecules representing each allele in the original sample. This copy number method can be combined with the evaluation of SNPs or other sequence variations to determine the chromosome copy number of born and unborn individuals; the discrimination and quantification of copies from loci which have short sequence variations, but in which PCR may amplifies from multiple target regions such as in carrier detection of Spinal Muscle Atrophy; determination of copy number of different sources of molecules from samples consisting of mixtures of different individual such as in detection of fetal aneuploidy from free floating DNA obtained from maternal plasma.

In an embodiment, the method as it pertains to a single target locus may comprise one or more of the following steps: (1) Designing a standard pair of oligomers for PCR amplification of a specific locus. (2) Adding, during synthesis, a sequence of specified bases with no or minimal complementarity to the target locus or genome to the 5' end of the one of the target specific oligomer. This sequence, termed the tail, is a known sequence, to be used for subsequent amplification, followed by a sequence of random nucleotides. These random nucleotides comprise the random region. The random region comprises a randomly generated sequence of nucleic acids that probabilistically differ between each probe molecule. Consequently, following synthesis, the tailed oligomer pool will consists of a collection of oligomers beginning with a known sequence followed by unknown sequence that differs between molecules, followed by the target specific sequence. (3) Performing one round of amplification (denaturation, annealing, extension) using only the tailed oligomer. (4) Adding exonuclease to the reaction, effectively stopping the PCR reaction, and incubating the reaction at the appropriate temperature to remove forward single stranded oligos that did not anneal to temple and extend to form a double stranded product. (5) Incubating the reaction at a high temperature to denature the exonuclease and eliminate its activity. (6) Adding to the reaction a new oligonucleotide that is complementary to tail of the oligomer used in the first reaction along with the other target specific oligomer to enable PCR amplification of the product generated in the first round of PCR. (7) Continuing amplification to generate enough product for downstream clonal sequencing. (8) Measuring the amplified PCR product by a multitude of methods, for example, clonal sequencing, to a sufficient number of bases to span the sequence.

In an embodiment, a method of the present disclosure involves targeting multiple loci in parallel or otherwise. Primers to different target loci can be generated independently and mixed to create multiplex PCR pools. In an embodiment, original samples can be divided into subpools and different loci can be targeted in each sub-pool before being recombined and sequenced. In an embodiment, the tagging step and a number of amplification cycles may be performed before the pool is subdivided to ensure efficient targeting of all targets before splitting, and improving subsequent amplification by continuing amplification using smaller sets of primers in subdivided pools.

One example of an application where this technology would be particularly useful is non-invasive prenatal aneuploidy diagnosis where the ratio of alleles at a given locus or a distribution of alleles at a number of loci can be used to help determine the number of copies of a chromosome present in a fetus. In this context, it is desirable to amplify the DNA present in the initial sample while maintaining the relative amounts of the various alleles. In some circumstances, especially in cases where there is a very small

## US 11,519,035 B2

119

amount of DNA, for example, fewer than 5,000 copies of the genome, fewer than 1,000 copies of the genome, fewer than 500 copies of the genome, and fewer than 100 copies of the genome, one can encounter a phenomenon called bottle-necking. This is where there are a small number of copies of 5 any given allele in the initial sample, and amplification biases can result in the amplified pool of DNA having significantly different ratios of those alleles than are in the initial mixture of DNA. By applying a unique or nearly unique set of barcodes to each strand of DNA before 10 standard PCR amplification, it is possible to exclude n-1 copies of DNA from a set of n identical molecules of sequenced DNA that originated from the same original

For example, imagine a heterozygous SNP in the genome 15 of an individual, and a mixture of DNA from the individual where ten molecules of each allele are present in the original sample of DNA. After amplification there may be 100,000 molecules of DNA corresponding to that locus. Due to stochastic processes, the ratio of DNA could be anywhere 20 from 1:2 to 2:1, however, since each of the original molecules was tagged with a unique tag, it would be possible to determine that the DNA in the amplified pool originated from exactly 10 molecules of DNA from each allele. This method would therefore give a more accurate measure of the 25 relative amounts of each allele than a method not using this approach. For methods where it is desirable for the relative amount of allele bias to be minimized, this method will provide more accurate data.

Association of the sequenced fragment to the target locus can be achieved in a number of ways. In an embodiment, a sequence of sufficient length is obtained from the targeted fragment to span the molecule barcode as well a sufficient number of unique bases corresponding to the target sequence to allow unambiguous identification of the target sequence to allow unambiguous identification of the target locus. In another embodiment, the molecular bar-coding primer that contains the randomly generated molecular barcode can also contain a locus specific barcode (locus barcode) that identifies the target to which it is to be associated. This locus barcode would be identical among all 40 molecular bar-coding primers for each individual target and hence all resulting amplicons, but different from all other targets. In an embodiment, the tagging method described herein may be combined with a one-sided nesting protocol.

In an embodiment, the design and generation of molecular 45 barcoding primers may be reduced to practice as follows: the molecular barcoding primers may consist of a sequence that is not complementary to the target sequence followed by random molecular barcode region followed by a target specific sequence. The sequence 5' of molecular barcode 50 may be used for subsequence PCR amplification and may comprise sequences useful in the conversion of the amplicon to a library for sequencing. The random molecular barcode sequence could be generated in a multitude of ways. The preferred method synthesize the molecule tagging primer in 55 such a way as to include all four bases to the reaction during synthesis of the barcode region. All or various combinations of bases may be specified using the IUPAC DNA ambiguity codes. In this manner the synthesized collection of molecules will contain a random mixture of sequences in the 60 molecular barcode region. The length of the barcode region will determine how many primers will contain unique barcodes. The number of unique sequences is related to the length of the barcode region as  $N^L$  where N is the number of bases, typically 4, and L is the length of the barcode. A 65 barcode of five bases can yield up to 1024 unique sequences; a barcode of eight bases can yield 65536 unique barcodes.

120

In an embodiment, the DNA can be measured by a sequencing method, where the sequence data represents the sequence of a single molecule. This can include methods in which single molecules are sequenced directly or methods in which single molecules are amplified to form clones detectable by the sequence instrument, but that still represent single molecules, herein called clonal sequencing.

Exemplary Methods and Reagents for Quantification of

Exemplary Methods and Reagents for Quantification of Amplification Products

Quantitation of specific nucleic acid sequences of interest is typically done by quantitative real-time PCR techniques such as TAQMAN (LIFE TECHNOLOGIES), INVADER probes (THIRD WAVE TECHNOLOGIES), and the like. Such techniques suffer from numerous shortcomings such as limited ability to achieve the simultaneous analysis of multiple sequences in parallel (multiplexation) and the ability to provide accurate quantitative data for only a narrow range of possible amplification cycles (e.g., when the logarithm of PCR amplification production quantity versus the number of cycles is in the linear range). DNA sequencing techniques, particularly high throughput next-generation sequencing techniques (often referred to as massively parallel sequencing techniques) such as those employed in MYSEQ (ILLU-MINA), HISEQ (ILLUMINA), ION TORRENT (LIFE TECHNOLOGIES), GENOME ANALYZER ILX (ILLU-MINA), GS FLEX+(ROCHE 454) etc., can be used for by quantitative measurements of the number of copies of sequence of interest present in sample, thereby providing quantitative information about the starting materials, e.g., copy number or transcription levels. High throughput genetic sequencers are amenable to the use of bar coding (i.e., sample tagging with distinctive nucleic acid sequences) so as to identify specific samples from individuals thereby permitting the simultaneous analysis of multiple samples in a single run of the DNA sequencer. The number of times a given region of the genome in a library preparation (or other nucleic preparation of interest) is sequenced (number of reads) will be proportional to the number of copies of that sequence in the genome of interest (or expression level in the case of cDNA containing preparations). However, the preparation and sequencing of genetic libraries (and similar genome derived preparations) can introduce numerous biases that interfere with obtaining an accurate quantitative reading for the nucleic acid sequence of interest. For example, different nucleic acid sequences can amplify with different efficiencies during nucleic amplification steps that take place during the genetic library preparation or sample preparation.

The problem with differential amplification efficiencies can be mitigated by using certain embodiments of the subject invention. The subject invention includes various methods and compositions that relate to the use of standards for inclusion in amplification processes that can be used to improve the accuracy of quantitation. The invention is of use in, among other areas, the detection of aneuploidy in a fetus by analyzing free floating fetal DNA in maternal blood, as described herein and as described, among other places, U.S. Pat. Nos. 8,008,018; 7,332,277; PCT Published Application WO 2012/078792A2; and PCT Published Application WO 2011/146632 A1, which are each herein incorporated by reference in its entirety Embodiments of the invention are also of use in the detection of aneuploidy in an in vitro generated embryos. Commercially significant aneuploidies that may be detected include aneuploidy of the human chromosomes 13, 18, 21, X and Y.

Embodiments of the invention may be used with either human or non-human nucleic acids, and may be applied to

## US 11,519,035 B2

12

both animal and plant derived nucleic acids. Embodiments of the invention may also be used to detect and/or quantitate alleles for other genetic disorders characterized by deletions or insertions. The deletion containing alleles can be detected in suspected carriers of the allele of interest.

One embodiment of the subject invention includes standards that are present in a known quantity (relative or absolute). For example, consider a genetic library made from a genetic source that is diploid for chromosome 8 (containing locus A) and triploid for chromosome 21 (con- 10 taining locus B). A genetic library can be produced from this sample that will contain sequences in quantities that are a function of the number of chromosomes present in the sample, e.g., 200 copies of locus A and 300 copies of locus B. However, if locus A amplifies much more efficiently than 15 locus B, after PCR there may be 60,000 copies of the A amplicon and 30,000 copies of the B amplicon, thus obscuring the true chromosomal copy number of the initial genomic sample when analysis by high throughput DNA sequencing (or other quantitative nucleic acid detection 20 techniques). To mitigate this problem a standard sequence for locus A is employed, wherein the standard sequence amplifies with essentially the same efficiency as locus A. Similarly, a standard sequence for locus B is created, wherein the standard sequence amplifies with the essentially 25 the same efficiency as locus B. A standard sequence of locus A and a standard sequence for locus B are added to the mixture prior to PCR (or other amplification techniques). These standard sequences are present in known quantities, either relative quantities or absolute quantities. Thus if a 1:1 mixture of standard sequence A and standard sequence B were added (prior to amplification) to the mixture in the previous example, 3000 copies of the standard A amplicon would be produced and 1000 copies of the standard B amplicon would be produced, showing that locus A is 35 amplified 3 times more efficiently than locus B, under the same set of conditions.

In various embodiments one or more selected regions of a genome containing a SNP (or other polymorphism) of interest can be specifically amplified and subsequently sequenced. This target specific amplification can take place during the formation of a genetic library for sequencing. The library can contain numerous targeted regions for amplification. In some embodiments at least 10; 100, 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 30,000; 40,000; 50,000; 75,000; or 100,000 regions of interest. Examples of such libraries are described herein and can be found in U.S. Patent Application No. 2012/0270212, filed Nov. 18, 2011, which is herein incorporated by reference in its entirety.

Many high throughput DNA sequencing techniques require the modification of the genetic starting material, e.g., the litigation of universal priming sites and/or barcodes, so as to form libraries to facilitate the clonal amplification of small nucleic acid fragments prior to performing subsequent 55 sequencing reactions. In some embodiments, one or more standard sequences are added during genetic library formation or added to a precursor component of a genetic library prior to amplification of the library. The standard sequences can be selected so as to mimic (yet be distinguishable based 60 on nucleotide base sequence) target genomic fragments to be prepared for sequencing by a high throughput genetic sequencing technique. In one embodiment, the standard sequence can be identical to the target genomic fragment excepting one, two, three, four to ten, or eleven to twenty 65 nucleotides. In some embodiments, when the target genetic sequence contains a SNP, the standard sequence can be

122

identical to the SNP excepting the nucleotide at the polymorphic base, which may be chosen to be one of the four nucleotides that is not observed at that location in nature. The standard sequences can be used in a highly multiplexed analysis of multiple target loci (such as polymorphic loci). Standard sequences can be added during the process of library formation (prior to amplification) in known quantities (relative or absolute) so as to provide a standard metric for greater accuracy in determining the amount of target sequence of interest in the sample of analysis. The combination of knowledge of the known quantities of the standard sequences used in conjunction with the knowledge of the ploidy level formation of library for sequencing formed from a genome of previously characterized ploidy level, e.g., known to be diploid for all autosomal chromosomes, can be used to calibrate the amplification properties of each standard sequence with respect to its corresponding target sequence and account for variations between batches of mixtures comprising multiple standard sequences. Given that it is often necessary to simultaneously analyze a large number of loci, it is useful to produce a mixture comprising a large set standard sequences. Embodiments of the invention include mixtures comprising multiple standard sequences. Ideally the amount of each standard sequence in the mixture is known with high precision. However, it is extremely difficult to achieve this ideal because as a practical matter there is a significant amount of variation in the quantity of each standard sequence in the mixture, particularly for mixtures comprising a large number of different synthetic oligonucleotides. This variation has numerous sources, e.g., variations in in vitro oligonucleotide synthesis reaction efficiencies between batch, inaccuracies in volume measurement, variations in pipetting, Furthermore, this variation can occur between different batches of that theoretically contain the exact same set of standard sequences in the exact same amounts. Accordingly, it is of interest to calibrate each batch of standard sequences independently. Batches of standard sequences can be calibrated against reference genomes of known chromosomal composition. Batched of standard sequences can be calibrated by sequencing the batch of standard sequences with minimal or no amplifications steps included in the sequencing protocol. Embodiments of the invention include calibrated mixtures of different standard sequences. Other embodiments of the invention include methods of calibrating mixtures of different standard sequences and calibrated mixtures of different standard sequences made by the subject methods.

Various embodiments of the subject mixtures of standard sequences and methods for using them can comprise at least 10; 100, 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 or more standards sequences, as well as various intermediate amounts. The number of the standard sequences can be the same as the number of target sequences selected for analysis during the generation of a targeted library for DNA sequencing. However, in some embodiments, it may be advantageous to use a lower number of standard sequences than the number of targeted regions in the library being constructed. It may be advantageous to use the lower number so as avoid coming up against the limits of the sequencing capacity of the high throughput DNA sequencer being employed. The number of standard sequences can be 50% or less than the number of targeted regions, 40% or less than the number of targeted regions, be 30% or less than the number of targeted regions, 20% or less than the number of targeted regions, be 10% or less than the number of targeted regions, 5% or less than the

## US 11,519,035 B2

12:

number of targeted regions, 1% or less than the number of targeted regions, as well as various intermediate values. For example, if a genetic library is created using 15,000 pairs of primers targeted to specific SNP containing loci, a suitable a mixture containing 1500 standard sequences corresponding to 1500 of the 15,000 targeted loci can be added prior to the amplification step of library constructions.

The amount of standard sequences added during library construction can vary considerably among different embodiments. In some embodiments, the amount of each standard 10 sequence can be approximately the same as the predicted amount of the target sequence present in the genomic material sample used for library preparation. In other embodiments, the amount of each standard sequence can be greater or less than the predicted amount of the target 15 sequence present in the genomic material sample used for library preparation. While the initial relative amounts of the target sequence and the standard sequence are not critical for the function of the invention, it is preferable that the amount be within the range 100 times greater to 100 times less than 20 the amount of the target sequence present in the genomic material sample used for library preparation. Excessive amounts of standard may use too much sequencing capacity of the DNA sequencer in a given run of the instrument. Using too low an amount of standard sequences will produce 25 insufficient data to aid in the analysis of variation in amplification efficiency.

The standard sequences may be selected to be very similar in nucleotide base sequence to the amplified regions of interest; preferably the standard sequence has the exact same 3 primer-binding sites as the analyzed genomic region, i.e., the "target sequence." The standard sequence must be distinguishable from the corresponding target sequence at a given locus. For the sake of convenience, this distinguishable region of the standard sequence will be referred to as a 3: "marker sequence." In some embodiments, the marker sequence region of the target sequences contains the polymorphic region, e.g., a SNP, and can be flanked on both sides by primer binding regions. The standard sequence may be selected to closely match the GC content of the correspond- 40 ing target sequence. In some embodiments, the primer binding regions of the standard sequence are flanked by universal priming sites. These universal priming sites are selected to match universal priming sites used in a genomic library for analysis. In other embodiments, the standard 4. sequences do not have universal priming sites and the universal priming sites are added during the creation of a library. Standard sequences are typically provided in single stranded form. A standard sequence is defined with respect to a corresponding target sequence and the sequence specific 50 reagents used to amplify the target sequence. In some embodiments, the target sequence contains the polymorphism of interest, e.g., a SNP, a deletion, or insertion, present in the nucleic acid sample for analysis. The standard sequence is a synthetic polynucleotide that is similar in 55 nucleotide base sequence to the target sequence, but is nonetheless distinguishable from the target sequence by virtue of at least one nucleotide base difference, thereby providing a mechanism for distinguishing amplicon sequences derived from the standard sequence form ampli- 60 con sequences derived from the target sequence. Standard sequences are selected so as to have essentially the same amplification properties as the corresponding target sequence when amplified with the same set of amplification reagents, e.g., PCR primers. In some embodiments, the 65 standard sequences can have the same primer sequence binding sites than the corresponding target sequences. In

124

other embodiments, the standard sequences can have a different primer sequence binding sites than the corresponding target sequences. In some embodiments, the standard sequences can be selected to produce amplicons that have the same length as the length of amplicons produced from the corresponding target sequences. In other embodiments, the standard sequences can be selected to produce amplicons that have the slightly different lengths than the length of amplicons produced from the corresponding target sequences.

After the amplification reactions have been completed, the library is sequenced on a high throughput DNA sequencer where individual molecule are clonally amplified and sequenced. The number of sequence reads for each allele of the target sequence is counted, also counted are the number of sequence reads for the standard sequence corresponding to the target sequence. The process is also carried out for at least one other pair of target sequences and corresponding standard sequences. Consider for example, locus A,  $X_{A1}$  reads for allele 1 of locus A are produced;  $X_{A2}$ reads for allele 2 of locus A are produced, and  $X_{AC}$  reads for standard sequence A are produced. The ratio of  $(X_{A1}$  plus  $X_{A2}$ ) to  $X_{AC}$  is determined for each locus of interest. As discussed earlier, the process can be performed on a reference genome, e.g., a genome that is known to be diploid for all chromosomes. The process can be repeated many times in order to provide a large number of read values so as to determine a mean number of reads and the standard deviation in the number of reads. The process is performed with a mixture comprising a large number of different standard sequences corresponding to different loci. By assuming that (1)  $X_{A1}$  plus  $X_{A2}$  corresponds to the known number of chromosome, e.g., 2 for the normal human female genome and (2) the standard sequences have similar amplification (and detectability) properties as their corresponding natural loci, the relative amounts of the different standard sequences in the multiplex standard mixture can be determined. The calibrated multiplex standard sequence mixture can then be used to adjust for the variability in amplification efficiency between the different loci in a multiplex amplification reac-

Other embodiments of the invention include methods and compositions for measuring the copy number of specific genes of interest, including duplications and mutant genes characterized by large deletions that would interfere with quantitation by sequencing. Sequencing would have problems detecting alleles having such deletions. Standard sequences included the amplification process can be used to reduce this problem.

In one embodiment of the invention the target sequence for analysis is a gene having a wild type (i.e. functional) form and a mutant form characterized by a deletion. Exemplary of such genes is SMN1, an allele having deletion being responsible for the genetic disease spinal muscular atrophy (SMA). It is of interest to detect an individual carrying the mutant form of the gene by means of high throughput genetic sequencing techniques. The application of such techniques to the detection of deletion mutations can be problematic because, among other reasons, the lack of sequences observed in sequencing (as opposed to detecting a simple point mutation or SNP). Such embodiments employ (1) a pair of amplification primers specific for the gene of interest, where in the amplification primers will amplify the gene of interest (or a portion thereof) and will not significantly amplify the mutant allele, (2) a standard sequence corresponding to the wild type allele of the gene of interest (i.e., a target sequence), but differing by at least one detect-

## US 11,519,035 B2

12:

able nucleotide base, (3) a pair of amplification primers specific for a second target sequence that serves as a reference sequence, and (4) a standard sequence corresponding to the reference sequence.

In one embodiment of the invention is provided a method 5 for measuring the number of copies of the gene of interest, where in the gene of interest has one meaning allele that comprises a deletion. The method can employ amplification reagent specific for the gene of interest, e.g., PCR primers, that are specific for the gene of interest by amplifying at least 10 a portion of the gene of interest, or the entire gene of interest, or a region adjacent to the gene of interest, while not amplifying the deletion comprising allele of the gene of interest. Additionally the subject method employs a standard sequence corresponding to the gene of interest, wherein the 15 standard sequence differs by at least one nucleotide base from the gene of interest (so that the sequence of the standard sequence can be readily distinguished from the naturally occurring gene of interest). Typically, the standard sequence will contain the same primer binding sites as the 20 gene of interest so as to minimize any amplification discrimination between the gene of interest and the standard sequence corresponding to the gene of interest. The reaction will also comprises amplification reagents specific for a reference sequence. The reference sequence is a sequence of 25 known (or at least assumed to be known) copy number in the genome to be analyzed. The reaction further comprises a standard sequence corresponding to the reference sequence. Typically, the standard sequence corresponding to the reference sequence will contain the same primer binding sites 30 as the reference sequence so as to minimize any amplification discrimination between the reference sequence and the standard sequence corresponding to the reference sequence. Exemplary PCR Conditions

If desired, any of the PCR conditions disclosed herein or any standard PCR conditions can be used to test a primer library to determine, e.g., the percent of primer dimers, percent of target amplicons, and percent of target loci that are amplified. If desired, standard methods can be used to optimize the reaction conditions to improve the performance of a primer library. Any of these PCR conditions may also be used in any of the methods of the invention to amplify target loci. It was determined that high ionic strength solutions can surprisingly be used for multiplex PCR. In some embodiments, monovalent cations are used to increase the 45 ionic strength to, e.g., help the primers bind the template.

In some embodiments, the reaction volume includes ethylenediaminetetraacetic acid (EDTA), magnesium, tetramethyl ammonium chloride (TMAC), or any combination thereof. In some embodiments, the concentration of TMAC 50 is between 20 and 80 mM, such as between 25 and 70 mM, 30 and 60 mM, 30 and 40 mM, 40 and 50 mM, 50 and 60 mM, or 60 and 70 mM, inclusive. While not meant to be bound to any particular theory, it is believed that TMAC binds to DNA, stabilizes duplexes, increases primer speci- 55 ficity, and/or equalizes the melting temperatures of different primers. In some embodiments, TMAC increases the uniformity in the amount of amplified products for the different targets. In some embodiments, the concentration of magnesium (such as magnesium from magnesium chloride) is 60 between 1 and 10 mM, such as between 1 and 8 mM, 1 and 5 mM, 1 and 3 mM, 3 and 5 mM, 3 and 6 mM, or 5 and 8 mM, inclusive.

In some embodiments, the concentration of available magnesium (the concentration of magnesium that is 65 assumed to be available for binding the polymerase and not bound to molecules other than the polymerase), such as the

126

magnesium that is not bound by phosphate groups on dNTPs, primers, or nucleic acid templates, or carboxylic acid groups on magnetic or other beads, if present) is between 0.5 to 10 mM, such as between 1 and 8 mM, 1 and 5 mM, 1 and 3 mM, 3 and 5 mM, 3 and 6 mM, 4 and 6 mM, or 5 and 8 mM, inclusive. The large number of primers used for multiplex PCR of a large number of targets may chelate a lot of the magnesium (2 phosphates in the primers chelate 1 magnesium). For example, if enough primers are used such that the concentration of phosphate from the primers is -9 mM, then the primers may reduce the effective magnesium concentration by ~4.5 mM. In some embodiments, EDTA is used to decrease the amount of magnesium available as a cofactor for the polymerase since high concentrations of magnesium can result in PCR errors, such as amplification of non-target loci. In some embodiments, the concentration of EDTA reduces the amount of available magnesium to between 1 and 5 mM (such as between 3 and 5 mM).

In some embodiments, the pH is between 7.5 and 8.5, such as between 7.5 and 8, 8 and 8.3, or 8.3 and 8.5, inclusive. In some embodiments, Tris is used at, for example, a concentration of between 10 and 100 mM, such as between 10 and 25 mM, 25 and 50 mM, 50 and 75 mM, or 25 and 75 mM, inclusive. In some embodiments, any of these concentrations of Tris are used at a pH between 7.5 and 8.5. In some embodiments, a combination of KCl and (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> is used, such as between 50 and 150 mM KCl and between 10 and 90 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, inclusive. In some embodiments, the concentration of KCl is between 0 and 30 mM, between 50 and 100 mM, or between 100 and 150 mM, inclusive. In some embodiments, the concentration of  $(NH_4)_2SO_4$  is between 10 and 50 mM, 50 and 90 mM, 10 and 20 mM, 20 and 40 mM, 40 mM and 60, or 60 mM and 80 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, inclusive. In some embodiments, the ammonium [NH<sub>4</sub><sup>+</sup>] concentration is between 0 and 160 mM, such as between 0 to 50, 50 to 100, or 100 to 160 mM, inclusive. In some embodiments, the sum of the potassium and ammonium concentration ([K+]+[NH<sub>4</sub>+]) is between 0 and 160 mM, such as between 0 to 25, 25 to 50, 50 to 150, 50 to 75, 75 to 100, 100 to 125, or 125 to 160 mM, inclusive. An exemplary buffer with  $[K+]+[NH_4+]=120$  mM is 20 mM KCl and 50 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. In some embodiments, the buffer includes 25 to 75 mM Tris, pH 7.2 to 8, 0 to 50 mM KCL, 10 to 80 mM ammonium sulfate, and 3 to 6 mM magnesium, inclusive. In some embodiments, the buffer includes 25 to 75 mM Tris pH 7 to 8.5, 3 to 6 mM MgC<sub>2</sub>, 10 to 50 mM KCl, and 20 to 80 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, inclusive. In some embodiments, 100 to 200 Units/m $\bar{L}$  of polymerase are used. In some embodiments, 100 mM KCl, 50 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 mM MgC<sub>2</sub>, 7.5 nM of each primer in the library, 50 mM TMAC, and 7 ul DNA template in a 20 ul final volume at pH 8.1 is used.

In some embodiments, a crowding agent is used, such as polyethylene glycol (PEG, such as PEG 8,000) or glycerol. In some embodiments, the amount of PEG (such as PEG 8,000) is between 0.1 to 20%, such as between 0.5 to 15%, 1 to 10%, 2 to 8%, or 4 to 8%, inclusive. In some embodiments, the amount of glycerol is between 0.1 to 20%, such as between 0.5 to 15%, 1 to 10%, 2 to 8%, or 4 to 8%, inclusive. In some embodiments, a crowding agent allows either a low polymerase concentration and/or a shorter annealing time to be used. In some embodiments, a crowding agent improves the uniformity of the DOR and/or reduces dropouts (undetected alleles). For example, at 8% PEG, and 50 U/mL polymerase, the uniformity was as good as 150 U/mL polymerase and no PEG. If the error rate

## US 11,519,035 B2

127

increases when PEG is included, a higher magnesium chloride concentration (such greater than or about 4, 5, 6, 7, 8, 9, or  $10~{\rm MgC_2}$ ) can be used to reduce or prevent the increase in error rate. Inclusion of 8% PEG 8,000 allowed successful multiplexing with an annealing time of only 1 minute at an 5 annealing temperature of  $63^{\circ}$  C.

In some embodiments, a polymerase with proof-reading activity, a polymerase without (or with negligible) proofreading activity, or a mixture of a polymerase with proofreading activity and a polymerase without (or with negligible) proof-reading activity is used. In some embodiments, a hot start polymerase, a non-hot start polymerase, or a mixture of a hot start polymerase and a non-hot start polymerase is used. In some embodiments, a HotStarTaq DNA polymerase is used (see, for example, QIAGEN cata- 15 log No. 203203, see, e.g., information available at the World Wide Web at giagen.com/us/products/catalog/assay-technologies/end-point-pcr-and-rt-pcr-reagents/hotstartaq-dnapolymerase/, which is hereby incorporated by reference in its entirety). In some embodiments, AmpliTaq Gold® DNA 20 Polymerase is used; it is a chemically modified form of AmpliTaq® DNA Polymerase requiring thermal activation for example, Applied Biosystems No.\_N8080241 see, e.g., information available at the World Wide Web at lifetechnoloies.com/order/catalog/product/ 25 N8080241, which is hereby incorporated by reference in its entirety). In some embodiments, KAPA Taq DNA Polymerase or KAPA Tag HotStart DNA Polymerase is used; they are based on the single-subunit, wild-type Taq DNA polymerase of the thermophilic bacterium Thermus aquati- 30 cus. KAPA Taq and KAPA Taq HotStart DNA Polymerase have 5'-3' polymerase and 5'-3' exonuclease activities, but no 3' to 5' exonuclease (proofreading) activity (see, for example, KAPA BIOSYSTEMS catalog No.\_BK1000 see, e.g., information available at the World Wide Web at kapa- 35 biosystems.com/product-applications/products/pcr-2/kapataq-pcr-kits/, which is hereby incorporated by reference in its entirety). In some embodiments, Pfu DNA polymerase is used; it is a highly thermostable DNA polymerase from the hyperthermophilic archaeum Pyrococcus furiosus. The 40 enzyme catalyzes the template-dependent polymerization of nucleotides into duplex DNA in the 5'-3' direction. Pfu DNA Polymerase also exhibits 3'-5' exonuclease (proofreading) activity that enables the polymerase to correct nucleotide incorporation errors. It has no 5'-3' exonuclease activity (see, 45 for example, Thermo Scientific catalog No.\_EP0501 see, e.g., information available at the World Wide Web at thermoscientificbio.com/pcr-enzymes-master-mixes-and-reagents/pfu-dna-polymerase/, which is hereby incorporated by reference in its entirety). In some embodiments Klentaql 50 is used; it is a Klenow-fragment analog of Taq DNA polymerase, it has no exonuclease or endonuclease activity (see, for example, DNA POLYMERASE TECHNOLOGY, Inc, St. Louis, Mo., catalog No.\_100 see, e.g., information available at the World Wide Web at klentaq.com/products/ 55 klentaq, which is hereby incorporated by reference in its entirety). In some embodiments, the polymerase is a PUSH-ION DNA polymerase, such as PHUSION High Fidelity DNA polymerase (M0530S, New England BioLabs, Inc.) or PHUSION Hot Start Flex DNA polymerase (M0535S, New 60 England BioLabs, Inc.; Frey and Suppman BioChemica. 2:34-35, 1995; Chester and Marshak Analytical Biochemistry. 209:284-290, 1993, which are each hereby incorporated by reference in its entirety). In some embodiments, the polymerase is a Q5® DNA Polymerase, such as Q5® 65 High-Fidelity DNA Polymerase (M0491S, New England BioLabs, Inc.) or Q5® Hot Start High-Fidelity DNA Poly128

merase (M0493S, New England BioLabs, Inc.). In some embodiments, the polymerase is a T4 DNA polymerase (M0203S, New England BioLabs, Inc.; Tabor and Struh. (1989). "DNA-Dependent DNA Polymerases," In Ausebel et al. (Ed.), *Current Protocols in Molecular Biology.* 3.5.10-3.5.12. New York: John Wiley & Sons, Inc., 1989; Sambrook et al. *Molecular Cloning: A Laboratory Manual.* (2nd ed.), 5.44-5.47. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989, which are each hereby incorporated by reference in its entirety).

In some embodiment, between 5 and 600 Units/mL (Units per 1 mL of reaction volume) of polymerase is used, such as between 5 to 100, 100 to 200, 200 to 300, 300 to 400, 400 to 500, or 500 to 600 Units/mL, inclusive. One unit is commonly defined as the amount of enzyme that will incorporate 15 nmol of dNTP into acid-insoluble material in 30 minutes at 75° C. Exemplary assay conditions for measuring unit activity include 1x THERMOPOL Reaction Buffer, 200  $\mu M$  dNTPs including [ $^3H$ ]-dTTP and 200  $\mu g/ml$ activated Calf Thymus DNA (see, e.g., information available at the world wide web at neb.com/products/m0267-taq-dnapolymerase-with-thermopol-buffer, which is hereby incorporated by reference in its entirety). 1× THERMOPOL® Reaction Buffer contains 20 mM Tris-HCl, 10 mM (NH<sub>4</sub>)<sub>2</sub> SO<sub>4</sub>, 10 mM KCl, 2 mM MgSO<sub>4</sub>, and 0.1% TRITON® X-100, pH 8.8.

In some embodiments, hot-start PCR is used to reduce or prevent polymerization prior to PCR thermocycling. Exemplary hot-start PCR methods include initial inhibition of the DNA polymerase, or physical separation of reaction components reaction until the reaction mixture reaches the higher temperatures. In some embodiments, the enzyme is spatially separated from the reaction mixture by wax that melts when the reaction reaches high temperature. In some embodiments, slow release of magnesium is used. DNA polymerase requires magnesium ions for activity, so the magnesium is chemically separated from the reaction by binding to a chemical compound, and is released into the solution only at high temperature. In some embodiments, non-covalent binding of an inhibitor is used. In this method a peptide, antibody, or aptamer are non-covalently bound to the enzyme at low temperature and inhibit its activity. After incubation at elevated temperature, the inhibitor is released and the reaction starts. In some embodiments, a coldsensitive Taq polymerase is used, such as a modified DNA polymerase with almost no activity at low temperature. In some embodiments, chemical modification is used. In this method, a molecule is covalently bound to the side chain of an amino acid in the active site of the DNA polymerase. The molecule is released from the enzyme by incubation of the reaction mixture at elevated temperature. Once the molecule is released, the enzyme is activated.

In some embodiments, the amount to template nucleic acids (such as an RNA or DNA sample) is between 20 and 5,000 ng, such as between 20 to 200, 200 to 400, 400 to 600, 600 to 1,000; 1,000 to 1,500; or 2,000 to 3,000 ng, inclusive.

In some embodiments QIAGEN Multiplex PCR Kit is used (QIAGEN catalog No. 206143; see, e.g., information available at the World Wide Web at giagen.com/products/catalog/assay-technologies/end-point-pcr-and-rt-pcr-reagents/qiagen-multiplex-pcr-kit, which is hereby incorporated by reference in its entirety). For 100×50 µl multiplex PCR reactions, the kit includes 2× QIAGEN Multiplex PCR Master Mix (providing a final concentration of 3 mM MgCl<sub>2</sub>, 3×0.85 ml), 5× Q-Solution (1×2.0 ml), and RNase-Free Water (2×1.7 ml). The QIAGEN Multiplex PCR Master Mix (MM) contains a combination of KCl and (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>

## US 11,519,035 B2

129

as well as the PCR additive, Factor MP, which increases the local concentration of primers at the template. Factor MP stabilizes specifically bound primers, allowing efficient primer extension by HotStarTaq DNA Polymerase. HotStarTaq DNA Polymerase is a modified form of Taq DNA polymerase and has no polymerase activity at ambient temperatures. In some embodiments, HotStarTaq DNA Polymerase is activated by a 15-minute incubation at 95° C. which can be incorporated into any existing thermal-cycler program.

In some embodiments, 1× QIAGEN MM final concentration (the recommended concentration), 7.5 nM of each primer in the library, 50 mM TMAC, and 7 ul DNA template in a 20 ul final volume is used. In some embodiments, the PCR thermocycling conditions include 95° C. for 10 minutes (hot start); 20 cycles of 96° C. for 30 seconds; 65° C. for 15 minutes; and 72° C. for 30 seconds; followed by 72° C. for 2 minutes (final extension); and then a 4° C. hold.

In some embodiments, 2× QIAGEN MM final concentration (twice the recommended concentration), 2 nM of 20 each primer in the library, 70 mM TMAC, and 7 ul DNA template in a 20 ul total volume is used. In some embodiments, up to 4 mM EDTA is also included. In some embodiments, the PCR thermocycling conditions include 95° C. for 10 minutes (hot start); 25 cycles of 96° C. for 30 25 seconds; 65° C. for 20 minutes; and 72° C. for 30 seconds); followed by 72° C. for 2 minutes (final extension); and then a 4° C. hold.

Another exemplary set of PCR thermocyling conditions includes 95° C. for 10 minutes, 15 cycles of 95° C. for 30 30 seconds, 65° C. for 1 minute, 60° C. for 5 minutes, 65° C. for 5 minutes and 72° C. for 30 seconds; and then 72° C. for 2 minutes. In some embodiments, this set of PCR thermocyling conditions is used with the following reaction conditions: 100 mM KCl, 50 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 mM MgCl<sub>2</sub>, 7.5 35 nM of each primer in the library, 50 mM TMAC, and 7 ul DNA template in a 20 ul final volume at pH 8.1.

Another exemplary set of conditions includes a seminested PCR approach. The first PCR reaction uses 20 ul a reaction volume with 2× QIAGEN MM final concentration, 40 1.875 nM of each primer in the library (outer forward and reverse primers), and DNA template.

Thermocycling parameters include 95° C. for 10 minutes; 25 cycles of 96° C. for 30 seconds, 65° C. for 1 minute, 58° C. for 6 minutes, 60° C. for 8 minutes, 65° C. for 4 minutes, 45 and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and then a 4° C. hold. Next, 2 ul of the resulting product, diluted 1:200, is as input in a second PCR reaction. This reaction uses a 10 ul reaction volume with 1× QIAGEN MM final concentration, 20 nM of each inner forward primer, and 1 50 uM of reverse primer tag. Thermocycling parameters include 95° C. for 10 minutes; 15 cycles of 95° C. for 30 seconds, 65° C. for 1 minute, 60° C. for 5 minutes, 65° C. for 5 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and then a 4° C. hold.

Any of the methods disclosed herein or any standard methods can be used to test a primer library to determine, e.g., the percent of primer dimers, percent of target amplicons, and percent of target loci that are amplified. In some embodiments, the PCR products are sequenced as described in Example 15 or using standard sequencing methods. In some embodiments, the percentage of primer dimers can be determined by measuring the number of sequencing reads from primer dimers, the percentage of amplified products that are target amplicons can be determined by measuring 65 the number of sequencing reads that map to target loci; the percent of target loci that are amplified can be determined by

130

measuring the number of target loci for which there are sequencing reads that map to the target loci; the number of copies of a particular amplified target loci can be determined based on the number of sequencing reads that map to that target loci (such as by comparing the number of sequencing reads compared to the sequences reads from a standard of known concentration or amount).

FIG. 45 contains data (such as percent mapped reads and error rate) from multiplex PCR with various buffers. In this figure, "1×MM" denotes 1× QIAGEN Master Mix (the recommended concentration) discussed above, and "2×MM" denotes 2× QIAGEN Master Mix (twice the recommended concentration). FIG. 45 also lists the components of buffer F-A (also called F-A Gold), F-B (also called F-B Gold), F-D, and F-J (also called F-B Qiagen or F-B Qia) as well as the amount and type of polymerase used to generate the data. FIG. 46 is a graph illustrating the uniformity in DOR for multiplex PCR with buffers from FIG. 45. FIG. 47 is a graph illustrating the normalized depth of read (DOR) for multiplex PCR with buffers from FIG. 45 with the DOR normalized to that of buffer 2×MM. Limit of Detection

As demonstrated by experiments provided in the Examples section, methods provided herein are capable of detecting an average allelic imbalance in a sample with a limit of detection or sensitivity of 0.45% AAI, which is the limit of detection for an uploidy of an illustrative method of the present invention. Similarly, in certain embodiments, methods provided herein are capable of detecting an average allelic imbalance in a sample of 0.45, 0.5, 0.6, 0.8, 0.8, 0.9, or 1.0%. That is, the test method is capable of detecting chromosomal aneuploidy in a sample down to an AAI of 0.45, 0.5, 0.6, 0.8, 0.8, 0.9, or 1.0%. As demonstrated by experiments provided in the Examples section, methods provided herein are capable of detecting the presence of an SNV in a sample for at least some SNVs, with a limit of detection or sensitivity of 0.2%, which is the limit of detection for at least some SNVs in one illustrative embodiment. Similarly, in certain embodiments, the method is capable of detecting an SNV with a frequency or SNV AAI of 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.8, 0.9, or 1.0%. That is, the test method is capable of detecting an SNV in a sample down to a limit of detection of 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.8, 0.9, or 1.0% of the total allele counts at the chromosomal locus of the SNV.

In some embodiments, a limit of detection of a mutation (such as an SNV or CNV) of a method of the invention is less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005%. In some embodiments, a limit of detection of a mutation (such as an SNV or CNV) of a method of the invention is between 15 to 0.005%, such as between 10 to 0.005%, 10 to 0.01%, 10 to 0.1%, 5 to 0.005%, 5 to 0.01%, 5 to 0.1%, 1 to 0.005%, 1 to 0.01%, 1 to 0.0

In some embodiments, a limit of detection is such that a mutation (such as an SNV or CNV) that is present in less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules with that locus in a sample (such as a sample of cfDNA or cfRNA) is detected (or is capable of being detected). For example, the mutation can be detected even if less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules that have that locus have that mutation in the locus (instead of, for example, a wild-type or non-mutated version of the locus or a different mutation at that locus). In some embodiments, a limit of detection is such that a mutation (such as an SNV or CNV) that is present in less than or equal to 10, 5, 2, 1,

## US 11,519,035 B2

13

0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample (such as a sample of cfDNA or cfRNA) is detected (or is capable of being detected). In some embodiments in which the CNV is a deletion, the deletion can be detected even if it is only present in less than 5 or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules that have a region of interest that may or may not contain the deletion in a sample. In some embodiments in which the CNV is a deletion, the deletion can be detected even if it is only present in less than or equal 10 to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample. In some embodiments in which the CNV is a duplication, the duplication can be detected even if the extra duplicated DNA or RNA that is present is less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 15 0.005% of the DNA or RNA molecules that have a region of interest that may or may not be duplicated in a sample in a sample. In some embodiments in which the CNV is a duplication, the duplication can be detected even if the extra duplicated DNA or RNA that is present is less than or equal 20 to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample. Experiment 23 provides exemplary methods for calculating the limit of detection. In some embodiments, the "LOD-zs5.0-mr5" method of Example 23 is used.

**Exemplary Nucleic Acid Samples** 

In some embodiments, the genetic sample may be prepared and/or purified. There are a number of standard procedures known in the art to accomplish such an end. In some embodiments, the sample may be centrifuged to separate various layers. In some embodiments, the DNA may be isolated using filtration. In some embodiments, the preparation of the DNA may involve amplification, separation, purification by chromatography, liquid separation, isolation, preferential enrichment, preferential amplification, targeted amplification, or any of a number of other techniques either known in the art or described herein.

In some embodiments, a method disclosed herein could be used in situations where there is a very small amount of DNA present, such as in in vitro fertilization, or in forensic 40 situations, where one or a few cells are available (typically less than ten cells, less than twenty cells or less than 40 cells.) In these embodiments, a method disclosed herein serves to make ploidy calls from a small amount of DNA that is not contaminated by other DNA, but where the ploidy 45 calling very difficult the small amount of DNA. In some embodiments, a method disclosed herein could be used in situations where the target DNA is contaminated with DNA of another individual, for example in maternal blood in the context of prenatal diagnosis, paternity testing, or products 50 of conception testing. Some other situations where these methods would be particularly advantageous would be in the case of cancer testing where only one or a small number of cells were present among a larger amount of normal cells. The genetic measurements used as part of these methods 55 could be made on any sample comprising DNA or RNA, for example but not limited to: blood, plasma, body fluids, urine, hair, tears, saliva, tissue, skin, fingernails, blastomeres, embryos, fetal cells, amniotic fluid, chorionic villus samples, feces, bile, lymph, cervical mucus, semen, or other 60 cells or materials comprising nucleic acids. In an embodiment, a method disclosed herein could be run with nucleic acid detection methods such as sequencing, microarrays, qPCR, digital PCR, or other methods used to measure nucleic acids. If for some reason it were found to be 65 desirable, the ratios of the allele count probabilities at a locus could be calculated, and the allele ratios could be used

132

to determine ploidy state in combination with some of the methods described herein, provided the methods are compatible. In some embodiments, a method disclosed herein involves calculating, on a computer, allele ratios at the plurality of polymorphic loci from the DNA measurements made on the processed samples. In some embodiments, a method disclosed herein involves calculating, on a computer, allele ratios at the plurality of polymorphic loci from the DNA measurements made on the processed samples along with any combination of other improvements described in this disclosure. Exemplary methods for isolating fetal cells, such as a single fetal cell are disclosed in U.S. Ser. No. 61/978,648, filed Apr. 11, 2014 and U.S. Ser. No. 61/984,546, filed Apr. 25, 2014. Fetal cells or fetal nucleic acids can be isolated from a pregnant mother using invasive (such as CVS or amniocentesis) or noninvasive methods (such as from a maternal blood sample).

In some embodiments, this method may be used to genotype a single cell, a small number of cells, two to five cells, six to ten cells, ten to twenty cells, twenty to fifty cells, fifty to one hundred cells, one hundred to one thousand cells. or a small amount of extracellular DNA, for example from one to ten picograms, from ten to one hundred pictograms,  $^{25}\,\,$  from one hundred pictograms to one nanogram, from one to ten nanograms, from ten to one hundred nanograms, from 30 to 500 nanograms, or from one hundred nanograms to one microgram. In some embodiments, nucleic acids (such as DNA and/or RNA) from less than 100, 75, 50, 40, 30, 20, 10, 8, 6, 4, 2, or 1 cell is amplified with any of the methods of the invention. In some embodiments, the nucleic acid sample includes less than 80, 60, 40, 20, or 10% of the nucleic acids (such as DNA and/or RNA) from a single cell. In some embodiments, in which a small number of cells (such as one cell) or a small amount of nucleic acids is used. nested PCR such as hemi-nested or semi-nested PCR is used and/or the number of PCR cycles is increased compare to that used for samples with a larger amount of cells or nucleic acids. In some embodiments, a large amount of cells or nucleic acids are used (such as in cases in which a larger amount is desired to improve performance of any of the methods of the invention. In some embodiments, a sample with at least 2, 5, 10, 15, 20, 30, 50, 100, or more cells (or DNA or RNA from such cells) is used in any of the methods of the invention. In some embodiments, at least 0.5, 1, 10, 25, 50, 100, 500, 1,000; or 5,000 ng of DNA or RNA is used.

In some embodiments, the cells in the sample are lysed prior to PCR. In some embodiments, the Arcturus PicoPure DNA extraction kit from Applied Biosystems is used. (Applied Biosystems cat. No. KIT0103, see, e.g., information available at the World Wide Web at lifetechnologies.com/order/catalog/product/KIT0103?ICID=search-product,

which is hereby incorporated by reference in its entirety). This kit contains Arcturus reconstitution buffer and Protease K. In some embodiments, the following cell lysis thermocycling protocol is used: 56° C. for 1 hour, 95° C. for 10 minutes, 25° C. for 15 minutes, and then a 4° C. hold.

In some embodiments, the nucleic acids are processed using the consecutive steps of end-repairing, dA-tailing, and adaptor ligating the nucleic acids. The consecutive steps exclude purifying the end-repaired products prior to the dA-tailing step and exclude purifying the dA-tailing products prior to the adaptor ligating step. The resulting products are amplified in any of the multiplex PCR methods of the invention. In some embodiments, the amplified products are then sequenced.

## US 11,519,035 B2

133

Exemplary Nucleic Acid Studies

The multiplex PCR methods of the invention can be used to increase the number of target loci that can be evaluated to measure the amount of one or more specific nucleic acid molecules of interest or of one or more types of nucleic acids. In some embodiments, there is a change in the total amount or concentration of one or more types of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic 10 RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA). In some embodiments, there is a change in the amount or concentration of one or more specific DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic 15 RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) molecules. In some embodiments, one allele is expressed more than another allele of a locus of interest. Exemplary miRNAs are short 20-22 nucleotide RNA mol- 20 ecules that regulate the expression of a gene. In some embodiments, there is a change in the transcriptome, such as a change in the identity or amount of one or more RNA

In some embodiments, an increase in the total amount or 25 concentration of cfDNA or cfRNA is associated with a disease such as cancer, or an increased risk for a disease such as cancer. In some embodiments, the total concentration of a type of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular 30 RNA, cytoplasmic RNA, coding cytoplasmic RNA, noncoding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) increases by at least 2, 3, 4, 5, 6, 7, 8, 9, 10-fold, or more compared to the total concentration of that type of DNA or RNA in healthy (such as non-cancerous) 3 subjects. In some embodiments, a total concentration of cfDNA between 75 to 100 ng/mL, 100 to 150 ng/mL, 150 to  $200~\rm ng/mL,~200$  to  $300~\rm ng/mL,~300$  to  $400~\rm ng/mgL,~400$  to  $600~\rm ng/mL,~600$  to  $800~\rm ng/mL,~800$  to  $1,000~\rm ng/mL,$  inclusive, or a total concentration of cfDNA of more than 100 ng, 40 mL, such as more than 200, 300, 400, 500, 600, 700, 800, 900, or 1,000 ng/mL is indicative of cancer, an increased risk for cancer, an increased risk of a tumor being malignant rather than benign, a decreased probably of the cancer going into remission, or a worse prognosis for the cancer. In some 4 embodiments, the amount of a type of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) having one 50 or more polymorphisms/mutations (such as deletions or duplications) associated with a disease such as cancer or an increased risk for a disease such as cancer is at least 2, 3, 4. 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25% of the total amount of that type of DNA or RNA. In some embodiments, 55 at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25%of the total amount of a type of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, 60 miRNA, mitochondrial RNA, rRNA, or tRNA) has a particular polymorphism or mutation (such as a deletion or duplication) associated with a disease such as cancer or an increased risk for a disease such as cancer.

Exemplary RNA Expression Studies

The multiplex PCR methods of the invention can be used to increase the number of target loci that can be evaluated 134

during gene expression profiling experiments. For example, the expression levels of thousands of genes can be simultaneously monitored to determine whether a person has a sequence (such as a polymorphism or other mutation) associated with a disease (such as cancer) or an increased risk of a disease. These methods can be used to identify sequences (such as polymorphisms or other mutations) associated with an increased or decreased risk for a disease such as cancer by comparing gene expression (such as the expression of particular mRNA alleles) in samples from patients with and without the disease. Additionally, the effect of particular treatments, diseases, or developmental stages on gene expression can be determined. Similarly, these methods can be used to identify genes whose expression is changed in response to pathogens or other organisms by comparing gene expression in infected and uninfected cells or tissues. In these methods the number of sequencing reads can be adjusted based on the frequency of the polymorphisms that are being analyzed such that sufficient reads are performed for the polymorphisms to be detected if they are present. In some embodiments, the polymorphisms or mutation is present at a higher frequency in subjects with a disease or disorder (such as cancer) than subjects without the disease or disorder (such as cancer). In some embodiments, the polymorphisms or mutation is indicative of cancer, such as a causative mutation.

In some embodiments, a sample containing RNA (such as mRNA) is amplified using a reverse transcriptase (RT) and the resulting DNA (such as cDNA) is then amplified using a DNA polymerase (PCR). The RT and PCR steps may be carried out sequentially in the same reaction volume or separately. Any of the primer libraries of the invention can be used in this reverse transcription polymerase chain reaction (RT-PCR) method. In various embodiments, the reverse transcription is performed using oligo-dT, random primers, a mixture of oligo-dT and random primers, or primers specific to the target loci. To avoid amplification of contaminating genomic DNA, primers for RT-PCR can be designed so that part of one primer hybridizes to the 3' end of one exon and the other part of the primer hybridizes to the 5' end of the adjacent exon. Such primers anneal to cDNA synthesized from spliced mRNAs, but not to genomic DNA. To detect amplification of contaminating DNA, RT-PCR primer pairs may be designed to flank a region that contains at least one intron. Products amplified from cDNA (no introns) are smaller than those amplified from genomic DNA (containing introns). Size difference in products is used to detect the presence of contaminating DNA. In some embodiments when only the mRNA sequence is known, primer annealing sites are chosen that are at least 300-400 base pairs apart since it is likely that fragments of this size from eukaryotic DNA contain splice junctions. Alternatively, the sample can be treated with DNase to degrade contaminating DNA.

Exemplary Methods for Paternity Testing

The multiplex PCR methods of the invention can be used to improve the accuracy of paternity testing since so many target loci can be analyzed at once (see, e.g, U.S. Publication No. 2012/0122701, filed Dec. 22, 2011, is which is hereby incorporated by reference in its entirety). For example, the multiplex PCR method can allow thousands of polymorphic loci (such as SNPs) to be analyzed for use in the PARENTAL SUPPORT algorithm described herein to determine whether an alleged father in is the biological father of a fetus. In some embodiments the method involves (i) simultaneously amplifying a plurality of polymorphic loci that includes at least 25; 50; 75; 100; 300; 500; 750; 1,000;

## US 11,519,035 B2

135

2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci on genetic material from the alleged father to produce a first set of amplified products; (ii) simultaneously amplifying the corresponding plurality of polymorphic loci on a mixed sample of DNA originating from a blood sample from the pregnant mother to produce a second set of amplified products; wherein the mixed sample of DNA comprises fetal DNA and maternal DNA; (iii) determining on a computer the probability that the alleged 10 father is the biological father of the fetus using genotypic measurements based on the first and second sets of amplified products; and (iv) establishing whether the alleged father is the biological father of the fetus using the determined probability that the alleged father is the biological father of 15 the fetus. In various embodiments, the method further includes simultaneously amplifying the corresponding plurality of polymorphic loci on genetic material from the mother to produce a third set of amplified products; wherein the probability that the alleged father is the biological father 20 of the fetus is determined using genotypic measurements based on the first, second, and third sets of amplified products.

Exemplary Methods for Embryo Characterization and Selection

The multiplex PCR methods of the invention can be used to improve the selection of embryos for in vitro fertilization by allowing thousands of target loci to be analyzed at once (see, e.g. U.S. Pub. No. 2011/0092763, filed May 27, 2008, filed Dec. 22, 2011, is which is hereby incorporated by reference in its entirety). For example, the multiplex PCR method can allow thousands of polymorphic loci (such as SNPs) to be analyzed for use in the PARENTAL SUPPORT algorithm described herein to select an embryo out of a set of embryos for in vitro fertilization

In some embodiments, the invention provides methods of estimating relative likelihoods that each embryo from a set of embryos will develop as desired. In some embodiments, the method involves contacting a sample from each embryo with a library of primers that simultaneously hybridize to at 40 least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci to produce a reaction mixture for each embryo, wherein the samples are each derived from one or more cells 45 from an embryo. In some embodiments, each reaction mixture is subjected to primer extension reaction conditions to produce amplified products. In some embodiments, the method includes determining on a computer one or more characteristics of at least one cell from each embryo based 50 on the amplified products; and estimating on a computer the relative likelihoods that each embryo will develop as desired, based on the one or more characteristics of the at least one cell for each embryo. In some embodiments, the method includes using an informatics based method to 55 determine the at least one characteristic, such as the PAREN-TAL SUPPORT algorithm described herein. In some embodiments, the characteristic includes a ploidy state. In some embodiments, the characteristic is selected from the group consisting of aneuploid, euploid, mosaic, nullsomy, 60 monosomy, uniparental disomy, trisomy, tetrasomy, a type of aneuploidy, unmatched copy error trisomy, matched copy error trisomy, maternal origin of aneuploidy, paternal origin of aneuploidy, a presence or absence of a disease-linked gene, a chromosomal identity of any aneuploid chromo- 65 some, an abnormal genetic condition, a deletion or duplication, a likelihood of a characteristic, and combinations

136

thereof. The characteristic may be associated with a chromosome taken from the group consisting of chromosome one, chromosome two, chromosome three, chromosome four, chromosome five, chromosome six, chromosome seven, chromosome eight, chromosome nine, chromosome ten, chromosome eleven, chromosome twelve, chromosome thirteen, chromosome fourteen, chromosome fifteen, chromosome sixteen, chromosome seventeen, chromosome eighteen, chromosome nineteen, chromosome twenty, chromosome twenty-one, chromosome twenty-two, X chromosome or Y chromosome, and combinations thereof.

Exemplary Prenatal Diagnostic Methods

The multipley PCR methods of the present

The multiplex PCR methods of the present invention can be used to improve prenatal diagnostic methods, such as the determination of the ploidy status of fetal chromosomes. Given that the large number of target loci that can be simultaneously amplified, more accurate determinations can be made.

In an embodiment, the present disclosure provides ex vivo methods for determining the ploidy status of a chromosome in a gestating fetus from genotypic data measured from a mixed sample of DNA (i.e., DNA from the mother of the fetus, and DNA from the fetus) and optionally from genotypic data measured from a sample of genetic material from 25 the mother and possibly also from the father, wherein the determining is done by using a joint distribution model to create a set of expected allele distributions for different possible fetal ploidy states given the parental genotypic data, and comparing the expected allelic distributions to the actual allelic distributions measured in the mixed sample, and choosing the ploidy state whose expected allelic distribution pattern most closely matches the observed allelic distribution pattern. In an embodiment, the mixed sample is derived from maternal blood, or maternal serum or plasma. In an embodiment, the mixed sample of DNA may be preferentially enriched at a target loci (e.g., plurality of polymorphic loci). In an embodiment, the preferential enrichment is done in a way that minimizes the allelic bias. In an embodiment, the present disclosure relates to a composition of DNA that has been preferentially enriched at a plurality of loci such that the allelic bias is low. In an embodiment, the allelic distribution(s) are measured by sequencing the DNA from the mixed sample. In an embodiment, the joint distribution model assumes that the alleles will be distributed in a binomial fashion. In an embodiment, the set of expected joint allele distributions are created for genetically linked loci while considering the extant recombination frequencies from various sources, for example, using data from the International HapMap Consortium.

In an embodiment, the present disclosure provides methods for non-invasive prenatal diagnosis (NPD), specifically, determining the aneuploidy status of a fetus by observing allele measurements at a plurality of polymorphic loci in genotypic data measured on DNA mixtures, where certain allele measurements are indicative of an aneuploid fetus, while other allele measurements are indicative of a euploid fetus. In an embodiment, the genotypic data is measured by sequencing DNA mixtures that were derived from maternal plasma. In an embodiment, the DNA sample may be preferentially enriched in molecules of DNA that correspond to the plurality of loci whose allele distributions are being calculated. In an embodiment a sample of DNA comprising only or almost only genetic material from the mother and possibly also a sample of DNA comprising only or almost only genetic material from the father are measured. In an embodiment, the genetic measurements of one or both parents along with the estimated fetal fraction are used to

## US 11,519,035 B2

137

create a plurality of expected allele distributions corresponding to different possible underlying genetic states of the fetus; the expected allele distributions may be termed hypotheses. In an embodiment, the maternal genetic data is not determined by measuring genetic material that is exclusively or almost exclusively maternal in nature, rather, it is estimated from the genetic measurements made on maternal plasma that comprises a mixture of maternal and fetal DNA. In some embodiments the hypotheses may comprise the ploidy of the fetus at one or more chromosomes, which 10 segments of which chromosomes in the fetus were inherited from which parents, and combinations thereof. In some embodiments, the ploidy state of the fetus is determined by comparing the observed allele measurements to the different hypotheses where at least some of the hypotheses corre- 15 spond to different ploidy states, and selecting the ploidy state that corresponds to the hypothesis that is most likely to be true given the observed allele measurements. In an embodiment, this method involves using allele measurement data from some or all measured SNPs, regardless of whether the 20 loci are homozygous or heterozygous, and therefore does not involve using alleles at loci that are only heterozygous. This method may not be appropriate for situations where the genetic data pertains to only one polymorphic locus. This method is particularly advantageous when the genetic data 25 comprises data for more than ten polymorphic loci for a target chromosome or more than twenty polymorphic loci. This method is especially advantageous when the genetic data comprises data for more than 50 polymorphic loci for a target chromosome, more than 100 polymorphic loci or 30 more than 200 polymorphic loci for a target chromosome. In some embodiments, the genetic data may comprise data for more than 500 polymorphic loci for a target chromosome, more than 1,000 polymorphic loci, more than 2,000 polymorphic loci, or more than 5,000 polymorphic loci for a 35 target chromosome.

In an embodiment, a method disclosed herein yields a quantitative measure of the number of independent observations of each allele at a polymorphic locus. This is unlike most methods such as microarrays or qualitative PCR which 40 provide information about the ratio of two alleles but do not quantify the number of independent observations of either allele. With methods that provide quantitative information regarding the number of independent observations, only the ratio is utilized in ploidy calculations, while the quantitative 45 information by itself is not useful. To illustrate the importance of retaining information about the number of independent observations consider the sample locus with two alleles, A and B. In a first experiment twenty A alleles and twenty B alleles are observed, in a second experiment 200 50 A alleles and 200 B alleles are observed. In both experiments the ratio (A/(A+B)) is equal to 0.5, however the second experiment conveys more information than the first about the certainty of the frequency of the A or B allele. Some methods by others involve averaging or summing allele 55 ratios (channel ratios) (i.e.  $x_i/y_i$ ) from individual allele and analyzes this ratio, either comparing it to a reference chromosome or using a rule pertaining to how this ratio is expected to behave in particular situations. No allele weighting is implied in such methods, where it is assumed that one 60 can ensure about the same amount of PCR product for each allele and that all the alleles should behave the same way. Such a method has a number of disadvantages, and more importantly, precludes the use a number of improvements that are described elsewhere in this disclosure.

In an embodiment, a method disclosed herein explicitly models the allele frequency distributions expected in disomy 138

as well as a plurality of allele frequency distributions that may be expected in cases of trisomy resulting from nondisjunction during meiosis I, nondisjunction during meiosis II, and/or nondisjunction during mitosis early in fetal development. To illustrate why this is important, imagine a case where there were no crossovers: nondisjunction during meiosis I would result a trisomy in which two different homologs were inherited from one parent; in contrast, nondisjunction during meiosis II or during mitosis early in fetal development would result in two copies of the same homolog from one parent. Each scenario would result in different expected allele frequencies at each polymorphic locus and also at all loci considered jointly, due to genetic linkage. Crossovers, which result in the exchange of genetic material between homologs, make the inheritance pattern more complex; in an embodiment, the instant method accommodates for this by using recombination rate information in addition to the physical distance between loci. In an embodiment, to enable improved distinction between meiosis I nondisjunction and meiosis II or mitotic nondisjunction the instant method incorporate into the model an increasing probability of crossover as the distance from the centromere increases. Meiosis II and mitotic nondisjunction can distinguished by the fact that mitotic nondisjunction typically results in identical or nearly identical copies of one homolog while the two homologs present following a meiosis II nondisjunction event often differ due to one or more crossovers during gametogenesis.

In some embodiments, a method disclosed herein involves comparing the observed allele measurements to theoretical hypotheses corresponding to possible fetal genetic aneuploidy, and does not involve a step of quantitating a ratio of alleles at a heterozygous locus. Where the number of loci is lower than about 20, the ploidy determination made using a method comprising quantitating a ratio of alleles at a heterozygous locus and a ploidy determination made using a method comprising comparing the observed allele measurements to theoretical allele distribution hypotheses corresponding to possible fetal genetic states may give a similar result. However, where the number of loci is above 50 these two methods is likely to give significantly different results; where the number of loci is above 400, above, 1.000 or above 2,000 these two methods are very likely to give results that are increasingly significantly different. These differences are due to the fact that a method that comprises quantitating a ratio of alleles at a heterozygous locus without measuring the magnitude of each allele independently and aggregating or averaging the ratios precludes the use of techniques including using a joint distribution model, performing a linkage analysis, using a binomial distribution model, and/or other advanced statistical techniques, whereas using a method comprising comparing the observed allele measurements to theoretical allele distribution hypotheses corresponding to possible fetal genetic states may use these techniques which can substantially increase the accuracy of the determination.

In an embodiment, a method disclosed herein involves determining whether the distribution of observed allele measurements is indicative of a euploid or an aneuploid fetus using a joint distribution model. The use of a joint distribution model is a different from and a significant improvement over methods that determine heterozygosity rates by treating polymorphic loci independently in that the resultant determinations are of significantly higher accuracy. Without being bound by any particular theory, it is believed that one reason they are of higher accuracy is that the joint distribution model takes into account the linkage between

## US 11,519,035 B2

139

SNPs, and likelihood of crossovers having occurred during the meiosis that gave rise to the gametes that formed the embryo that grew into the fetus. The purpose of using the concept of linkage when creating the expected distribution of allele measurements for one or more hypotheses is that it allows the creation of expected allele measurements distributions that correspond to reality considerably better than when linkage is not used. For example, imagine that there are two SNPs, 1 and 2 located nearby one another, and the mother is A at SNP 1 and A at SNP 2 on one homolog, and 10 B at SNP 1 and B at SNP 2 on homolog two. If the father is A for both SNPs on both homologs, and a B is measured for the fetus SNP 1, this indicates that homolog two has been inherited by the fetus, and therefore that there is a much higher likelihood of a B being present on the fetus at SNP 15 2. A model that takes into account linkage would predict this, while a model that does not take linkage into account would not. Alternately, if a mother was AB at SNP 1 and AB at nearby SNP 2, then two hypotheses corresponding to maternal trisomy at that location could be used—one involving a 20 matching copy error (nondisjunction in meiosis II or mitosis in early fetal development), and one involving an unmatching copy error (nondisjunction in meiosis I). In the case of a matching copy error trisomy, if the fetus inherited an AA from the mother at SNP 1, then the fetus is much more likely 25 to inherit either an AA or BB from the mother at SNP 2, but not AB. In the case of an unmatching copy error, the fetus would inherit an AB from the mother at both SNPs. The allele distribution hypotheses made by a ploidy calling method that takes into account linkage would make these 30 predictions, and therefore correspond to the actual allele measurements to a considerably greater extent than a ploidy calling method that did not take into account linkage. Note that a linkage approach is not possible when using a method that relies on calculating allele ratios and aggregating those 35 allele ratios.

One reason that it is believed that ploidy determinations that use a method that comprises comparing the observed allele measurements to theoretical hypotheses corresponding to possible fetal genetic states are of higher accuracy is 40 that when sequencing is used to measure the alleles, this method can glean more information from data from alleles where the total number of reads is low than other methods; for example, a method that relies on calculating and aggregating allele ratios would produce disproportionately 4. weighted stochastic noise. For example, imagine a case that involved measuring the alleles using sequencing, and where there was a set of loci where only five sequence reads were detected for each locus. In an embodiment, for each of the alleles, the data may be compared to the hypothesized allele 50 distribution, and weighted according to the number of sequence reads; therefore the data from these measurements would be appropriately weighted and incorporated into the overall determination. This is in contrast to a method that involved quantitating a ratio of alleles at a heterozygous 55 locus, as this method could only calculate ratios of 0%, 20%, 40%, 60%, 80% or 100% as the possible allele ratios; none of these may be close to expected allele ratios. In this latter case, the calculated allele rations would either have to be discarded due to insufficient reads or else would have 60 disproportionate weighting and introduce stochastic noise into the determination, thereby decreasing the accuracy of the determination. In an embodiment, the individual allele measurements may be treated as independent measurements, where the relationship between measurements made on 65 alleles at the same locus is no different from the relationship between measurements made on alleles at different loci.

140

In an embodiment, a method disclosed herein involves determining whether the distribution of observed allele measurements is indicative of a euploid or an aneuploid fetus without comparing any metrics to observed allele measurements on a reference chromosome that is expected to be disomic (termed the RC method). This is a significant improvement over methods, such as methods using shotgun sequencing which detect aneuploidy by evaluating the proportion of randomly sequenced fragments from a suspect chromosomes relative to one or more presumed disomic reference chromosome. This RC method yields incorrect results if the presumed disomic reference chromosome is not actually disomic. This can occur in cases where aneuploidy is more substantial than trisomy of a single chromosome or where the fetus is triploid and all autosomes are trisomic. In the case of a female triploid (69, XXX) fetus there are in fact no disomic chromosomes at all. The method described herein does not require a reference chromosome and would be able to correctly identify trisomic chromosomes in a female triploid fetus. For each chromosome, hypothesis, child fraction and noise level, a joint distribution model may be fit, without any of: reference chromosome data, an overall child fraction estimate, or a fixed reference hypothesis.

In an embodiment, a method disclosed herein demonstrates how observing allele distributions at polymorphic loci can be used to determine the ploidy state of a fetus with greater accuracy than methods in the prior art. In an embodiment, the method uses the targeted sequencing to obtain mixed maternal-fetal genotypes and optionally mother and/ or father genotypes at a plurality of SNPs to first establish the various expected allele frequency distributions under the different hypotheses, and then observing the quantitative allele information obtained on the maternal-fetal mixture and evaluating which hypothesis fits the data best, where the genetic state corresponding to the hypothesis with the best fit to the data is called as the correct genetic state. In an embodiment, a method disclosed herein also uses the degree of fit to generate a confidence that the called genetic state is the correct genetic state. In an embodiment, a method disclosed herein involves using algorithms that analyze the distribution of alleles found for loci that have different parental contexts, and comparing the observed allele distributions to the expected allele distributions for different ploidy states for the different parental contexts (different parental genotypic patterns). This is different from and an improvement over methods that do not use methods that enable the estimation of the number of independent instances of each allele at each locus in a mixed maternalfetal sample. In an embodiment, a method disclosed herein involves determining whether the distribution of observed allele measurements is indicative of a euploid or an aneuploid fetus using observed allelic distributions measured at loci where the mother is heterozygous. This is different from and an improvement over methods that do not use observed allelic distributions at loci where the mother is heterozygous because, in cases where the DNA is not preferentially enriched or is preferentially enriched for loci that are not known to be highly informative for that particular target individual, it allows the use of about twice as much genetic measurement data from a set of sequence data in the ploidy determination, resulting in a more accurate determination.

In an embodiment, a method disclosed herein uses a joint distribution model that assumes that the allele frequencies at each locus are multinomial (and thus binomial when SNPs are biallelic) in nature. In some embodiments the joint distribution model uses beta-binomial distributions. When using a measuring technique, such as sequencing, provides

## US 11,519,035 B2

141

a quantitative measure for each allele present at each locus, binomial model can be applied to each locus and the degree underlying allele frequencies and the confidence in that frequency can be ascertained. With methods known in the art that generate ploidy calls from allele ratios, or methods in 5 which quantitative allele information is discarded, the certainty in the observed ratio cannot be ascertained. The instant method is different from and an improvement over methods that calculate allele ratios and aggregate those ratios to make a ploidy call, since any method that involves 10 calculating an allele ratio at a particular locus, and then aggregating those ratios, necessarily assumes that the measured intensities or counts that are indicative of the amount of DNA from any given allele or locus will be distributed in a Gaussian fashion. The method disclosed herein does not 15 involve calculating allele ratios. In some embodiments, a method disclosed herein may involve incorporating the number of observations of each allele at a plurality of loci into a model. In some embodiments, a method disclosed herein may involve calculating the expected distributions 20 themselves, allowing the use of a joint binomial distribution model which may be more accurate than any model that assumes a Gaussian distribution of allele measurements. The likelihood that the binomial distribution model is significantly more accurate than the Gaussian distribution 25 increases as the number of loci increases. For example, when fewer than 20 loci are interrogated, the likelihood that the binomial distribution model is significantly better is low. However, when more than 100, or especially more than 400, or especially more than 1,000, or especially more than 2,000 30 loci are used, the binomial distribution model will have a very high likelihood of being significantly more accurate than the Gaussian distribution model, thereby resulting in a more accurate ploidy determination. The likelihood that the binomial distribution model is significantly more accurate 3 than the Gaussian distribution also increases as the number of observations at each locus increases. For example, when fewer than 10 distinct sequences are observed at each locus are observed, the likelihood that the binomial distribution model is significantly better is low. However, when more 40 than 50 sequence reads, or especially more than 100 sequence reads, or especially more than 200 sequence reads, or especially more than 300 sequence reads are used for each locus, the binomial distribution model will have a very high likelihood of being significantly more accurate than the 45 Gaussian distribution model, thereby resulting in a more accurate ploidy determination.

In an embodiment, a method disclosed herein uses sequencing to measure the number of instances of each allele at each locus in a DNA sample. Each sequencing read 50 may be mapped to a specific locus and treated as a binary sequence read; alternately, the probability of the identity of the read and/or the mapping may be incorporated as part of the sequence read, resulting in a probabilistic sequence read, that is, the probable whole or fractional number of sequence 55 reads that map to a given loci. Using the binary counts or probability of counts it is possible to use a binomial distribution for each set of measurements, allowing a confidence interval to be calculated around the number of counts. This ability to use the binomial distribution allows for more 60 accurate ploidy estimations and more precise confidence intervals to be calculated. This is different from and an improvement over methods that use intensities to measure the amount of an allele present, for example methods that use microarrays, or methods that make measurements using 65 fluorescence readers to measure the intensity of fluorescently tagged DNA in electrophoretic bands.

142

In an embodiment, a method disclosed herein uses aspects of the present set of data to determine parameters for the estimated allele frequency distribution for that set of data. This is an improvement over methods that utilize training set of data or prior sets of data to set parameters for the present expected allele frequency distributions, or possibly expected allele ratios. This is because there are different sets of conditions involved in the collection and measurement of every genetic sample, and thus a method that uses data from the instant set of data to determine the parameters for the joint distribution model that is to be used in the ploidy determination for that sample will tend to be more accurate.

In an embodiment, a method disclosed herein involves determining whether the distribution of observed allele measurements is indicative of a euploid or an aneuploid fetus using a maximum likelihood technique. The use of a maximum likelihood technique is different from and a significant improvement over methods that use single hypothesis rejection technique in that the resultant determinations will be made with significantly higher accuracy. One reason is that single hypothesis rejection techniques set cut off thresholds based on only one measurement distribution rather than two, meaning that the thresholds are usually not optimal. Another reason is that the maximum likelihood technique allows the optimization of the cut off threshold for each individual sample instead of determining a cut off threshold to be used for all samples regardless of the particular characteristics of each individual sample. Another reason is that the use of a maximum likelihood technique allows the calculation of a confidence for each ploidy call. The ability to make a confidence calculation for each call allows a practitioner to know which calls are accurate, and which are more likely to be wrong. In some embodiments, a wide variety of methods may be combined with a maximum likelihood estimation technique to enhance the accuracy of the ploidy calls. In an embodiment, the maximum likelihood technique may be used in combination with the method described in U.S. Pat. No. 7,888,017. In an embodiment, the maximum likelihood technique may be used in combination with the method of using targeted PCR amplification to amplify the DNA in the mixed sample followed by sequencing and analysis using a read counting method such as used by TANDEM DIAGNOSTICS, as presented at the International Congress of Human Genetics 2011, in Montreal in October 2011. In an embodiment, a method disclosed herein involves estimating the fetal fraction of DNA in the mixed sample and using that estimation to calculate both the ploidy call and the confidence of the ploidy call. Note that this is both different and distinct from methods that use estimated fetal fraction as a screen for sufficient fetal fraction, followed by a ploidy call made using a single hypothesis rejection technique that does not take into account the fetal fraction nor does it produce a confidence calculation for the call.

In an embodiment, a method disclosed herein takes into account the tendency for the data to be noisy and contain errors by attaching a probability to each measurement. The use of maximum likelihood techniques to choose the correct hypothesis from the set of hypotheses that were made using the measurement data with attached probabilistic estimates makes it more likely that the incorrect measurements will be discounted, and the correct measurements will be used in the calculations that lead to the ploidy call. To be more precise, this method systematically reduces the influence of data that is incorrectly measured on the ploidy determination. This is an improvement over methods where all data is assumed to be equally correct or methods where outlying data is arbi-

## US 11,519,035 B2

143

trarily excluded from calculations leading to a ploidy call. Existing methods using channel ratio measurements claim to extend the method to multiple SNPs by averaging individual SNP channel ratios. Not weighting individual SNPs by expected measurement variance based on the SNP quality and observed depth of read reduces the accuracy of the resulting statistic, resulting in a reduction of the accuracy of the ploidy call significantly, especially in borderline cases.

In an embodiment, a method disclosed herein does not presuppose the knowledge of which SNPs or other polymorphic loci are heterozygous on the fetus. This method allows a ploidy call to be made in cases where paternal genotypic information is not available. This is an improvement over methods where the knowledge of which SNPs are heterozygous must be known ahead of time in order to 15 appropriately select loci to target, or to interpret the genetic measurements made on the mixed fetal/maternal DNA sample.

The methods described herein are particularly advantageous when used on samples where a small amount of DNA 20 is available, or where the percent of fetal DNA is low. This is due to the correspondingly higher allele dropout rate that occurs when only a small amount of DNA is available and/or the correspondingly higher fetal allele dropout rate when the percent of fetal DNA is low in a mixed sample of fetal and 25 maternal DNA. A high allele dropout rate, meaning that a large percentage of the alleles were not measured for the target individual, results in poorly accurate fetal fractions calculations, and poorly accurate ploidy determinations. Since methods disclosed herein may use a joint distribution 30 model that takes into account the linkage in inheritance patterns between SNPs, significantly more accurate ploidy determinations may be made. The methods described herein allow for an accurate ploidy determination to be made when the percent of molecules of DNA that are fetal in the mixture 3: is less than 40%, less than 30%, less than 20%, less than 10%, less than 8%, and even less than 6%.

In an embodiment, it is possible to determine the ploidy state of an individual based on measurements when that individual's DNA is mixed with DNA of a related indi- 40 vidual. In an embodiment, the mixture of DNA is the free floating DNA found in maternal plasma, which may include DNA from the mother, with known karyotype and known genotype, and which may be mixed with DNA of the fetus, with unknown karyotype and unknown genotype. It is 45 possible to use the known genotypic information from one or both parents to predict a plurality of potential genetic states of the DNA in the mixed sample for different ploidy states, different chromosome contributions from each parent to the fetus, and optionally, different fetal DNA fractions in 50 the mixture. Each potential composition may be referred to as a hypothesis. The ploidy state of the fetus can then be determined by looking at the actual measurements, and determining which potential compositions are most likely given the observed data.

Further discussion of the points above may be found elsewhere in this document.

Non-Invasive Prenatal Diagnosis (NPD)

The process of non-invasive prenatal diagnosis involves a number of steps. Some of the steps may include: (1) obtaining the genetic material from the fetus; (2) enriching the genetic material of the fetus that may be in a mixed sample, ex vivo; (3) amplifying the genetic material, ex vivo; (4) preferentially enriching specific loci in the genetic material, ex vivo; (5) measuring the genetic material, ex vivo; and (6) 65 analyzing the genotypic data, on a computer, and ex vivo. Methods to reduce to practice these six and other relevant

144

steps are described herein. At least some of the method steps are not directly applied on the body. In an embodiment, the present disclosure relates to methods of treatment and diagnosis applied to tissue and other biological materials isolated and separated from the body. At least some of the method steps are executed on a computer.

Some embodiments of the present disclosure allow a clinician to determine the genetic state of a fetus that is gestating in a mother in a non-invasive manner such that the health of the baby is not put at risk by the collection of the genetic material of the fetus, and that the mother is not required to undergo an invasive procedure. Moreover, in certain aspects, the present disclosure allows the fetal genetic state to be determined with high accuracy, significantly greater accuracy than, for example, the non-invasive maternal serum analyte based screens, such as the triple test, that are in wide use in prenatal care.

The high accuracy of the methods disclosed herein is a result of an informatics approach to analysis of the genotype data, as described herein. Modern technological advances have resulted in the ability to measure large amounts of genetic information from a genetic sample using such methods as high throughput sequencing and genotyping arrays. The methods disclosed herein allow a clinician to take greater advantage of the large amounts of data available, and make a more accurate diagnosis of the fetal genetic state. The details of a number of embodiments are given below. Different embodiments may involve different combinations of the aforementioned steps. Various combinations of the different embodiments of the different steps may be used interchangeably.

In an embodiment, a blood sample is taken from a pregnant mother, and the free floating DNA in the plasma of the mother's blood, which contains a mixture of both DNA of maternal origin, and DNA of fetal origin, is isolated and used to determine the ploidy status of the fetus. In an embodiment, a method disclosed herein involves preferential enrichment of those DNA sequences in a mixture of DNA that correspond to polymorphic alleles in a way that the allele ratios and/or allele distributions remain mostly consistent upon enrichment. In an embodiment, a method disclosed herein involves the highly efficient targeted PCR based amplification such that a very high percentage of the resulting molecules correspond to targeted loci. In an embodiment, a method disclosed herein involves sequencing a mixture of DNA that contains both DNA of maternal origin, and DNA of fetal origin. In an embodiment, a method disclosed herein involves using measured allele distributions to determine the ploidy state of a fetus that is gestating in a mother. In an embodiment, a method disclosed herein involves reporting the determined ploidy state to a clinician. In an embodiment, a method disclosed herein involves taking a clinical action, for example, performing follow up invasive testing such as chorionic villus sampling or amniocentesis, preparing for the birth of a trisomic individual or an elective termination of a trisomic fetus.

This application makes reference to U.S. Utility application Ser. No. 11/603,406, filed Nov. 28, 2006 (US Publication No.: 20070184467); U.S. Utility application Ser. No. 12/076,348, filed Mar. 17, 2008 (US Publication No.: 20080243398); PCT Application Serial No. PCT/US09/52730, filed Aug. 4, 2009 (PCT Publication No.: WO/2010/017214); PCT Application Serial No. PCT/US10/050824, filed Sep. 30, 2010 (PCT Publication No.: WO/2011/041485), U.S. Utility application Ser. No. 13/110,685, filed May 18, 2011, and PCT Application Serial No. PCT/12/58578, filed Oct. 3, 2012, which are each herein incorpo-

Case: 24-1324 Document: 42-1 Page: 195 Filed: 03/18/2024

## US 11,519,035 B2

145

rated by reference in its entirety. Some of the vocabulary used in this filing may have its antecedents in these references. Some of the concepts described herein may be better understood in light of the concepts found in these references. Screening Maternal Blood Comprising Free Floating Fetal 5 DNA

The methods described herein may be used to help determine the genotype of a child, fetus, or other target individual where the genetic material of the target is found in the presence of a quantity of other genetic material. In 10 some embodiments the genotype may refer to the ploidy state of one or a plurality of chromosomes, it may refer to one or a plurality of disease linked alleles, or some combination thereof. In this disclosure, the discussion focuses on determining the genetic state of a fetus where the fetal DNA 15 is found in maternal blood, but this example is not meant to limit to possible contexts that this method may be applied to. In addition, the method may be applicable in cases where the amount of target DNA is in any proportion with the nontarget DNA; for example, the target DNA could make up 20 anywhere between 0.000001 and 99.999999% of the DNA present. In addition, the non-target DNA does not necessarily need to be from one individual, or even from a related individual, as long as genetic data from some or all of the ment, a method disclosed herein can be used to determine genotypic data of a fetus from maternal blood that contains fetal DNA. It may also be used in a case where there are multiple fetuses in the uterus of a pregnant woman, or where other contaminating DNA may be present in the sample, for 30 example from other already born siblings.

This technique may make use of the phenomenon of fetal blood cells gaining access to maternal circulation through the placental villi. Ordinarily, only a very small number of fetal cells enter the maternal circulation in this fashion (not 3: enough to produce a positive Kleihauer-Betke test for fetalmaternal hemorrhage). The fetal cells can be sorted out and analyzed by a variety of techniques to look for particular DNA sequences, but without the risks that invasive procedures inherently have. This technique may also make use of 40 the phenomenon of free floating fetal DNA gaining access to maternal circulation by DNA release following apoptosis of placental tissue where the placental tissue in question contains DNA of the same genotype as the fetus. The free floating DNA found in maternal plasma has been shown to 45 contain fetal DNA in proportions as high as 30-40% fetal DNA.

In an embodiment, blood may be drawn from a pregnant woman. Research has shown that maternal blood may contain a small amount of free floating DNA from the fetus, in 50 addition to free floating DNA of maternal origin. In addition, there also may be enucleated fetal blood cells comprising DNA of fetal origin, in addition to many blood cells of maternal origin, which typically do not contain nuclear DNA. There are many methods know in the art to isolate 55 fetal DNA, or create fractions enriched in fetal DNA. For example, chromatography has been show to create certain fractions that are enriched in fetal DNA.

Once the sample of maternal blood, plasma, or other fluid, drawn in a relatively non-invasive manner, and that contains 60 an amount of fetal DNA, either cellular or free floating, either enriched in its proportion to the maternal DNA, or in its original ratio, is in hand, one may genotype the DNA found in said sample. In some embodiments, the blood may be drawn using a needle to withdraw blood from a vein, for 65 example, the basilica vein. The method described herein can be used to determine genotypic data of the fetus. For

example, it can be used to determine the ploidy state at one or more chromosomes, it can be used to determine the identity of one or a set of SNPs, including insertions, deletions, and translocations. It can be used to determine one or more haplotypes, including the parent of origin of one or more genotypic features.

Note that this method will work with any nucleic acids that can be used for any genotyping and/or sequencing methods, such as the ILLUMINA INFINIUM ARRAY platform, AFFYMETRIX GENECHIP, ILLUMINA GENOME ANALYZER, or LIFE TECHNOLGIES' SOLID SYSTEM. This includes extracted free-floating DNA from plasma or amplifications (e.g. whole genome amplification, PCR) of the same; genomic DNA from other cell types (e.g. human lymphocytes from whole blood) or amplifications of the same. For preparation of the DNA, any extraction or purification method that generates genomic DNA suitable for the one of these platforms will work as well. This method could work equally well with samples of RNA. In an embodiment, storage of the samples may be done in a way that will minimize degradation (e.g. below freezing, at about -20 C, or at a lower temperature).

Parental Support

Some embodiments may be used in combination with the relevant non-target individual(s) is known. In an embodi- 25 PARENTAL SUPPORT<sup>TM</sup> (PS) method, embodiments of which are described in U.S. application Ser. No. 11/603,406 (US Publication No.: 20070184467), U.S. application Ser. No. 12/076,348 (US Publication No.: 20080243398), U.S. application Ser. No. 13/110,685, PCT Application PCT/ US09/52730 (PCT Publication No.: WO/2010/017214), and PCT Application No. PCT/US10/050824 (PCT Publication No.: WO/2011/041485) which are incorporated herein by reference in their entirety. PARENTAL SUPPORT<sup>TM</sup> is an informatics based approach that can be used to analyze genetic data. In some embodiments, the methods disclosed herein may be considered as part of the PARENTAL SUP-PORT™ method. In some embodiments, The PARENTAL SUPPORT<sup>TM</sup> method is a collection of methods that may be used to determine the genetic data of a target individual, with high accuracy, of one or a small number of cells from that individual, or of a mixture of DNA consisting of DNA from the target individual and DNA from one or a plurality of other individuals, specifically to determine disease-related alleles, other alleles of interest, and/or the ploidy state of one or a plurality of chromosomes in the target individual. PARENTAL SUPPORT™ may refer to any of these methods. PARENTAL SUPPORT™ is an example of an informatics based method. Exemplary embodiments of the PARENTAL SUPPORT<sup>TM</sup> method are illustrated in FIGS. **29-31**G and described in Example 19.

The PARENTAL SUPPORT<sup>TM</sup> method makes use of known parental genetic data, i.e. haplotypic and/or diploid genetic data of the mother and/or the father, together with the knowledge of the mechanism of meiosis and the imperfect measurement of the target DNA, and possibly of one or more related individuals, along with population based crossover frequencies, in order to reconstruct, in silico, the genotype at a plurality of alleles, and/or the ploidy state of an embryo or of any target cell(s), and the target DNA at the location of key loci with a high degree of confidence. The PARENTAL SUPPORT™ method can reconstruct not only single nucleotide polymorphisms (SNPs) that were measured poorly, but also insertions and deletions, and SNPs or whole regions of DNA that were not measured at all. Furthermore, the PARENTAL SUPPORT<sup>TM</sup> method can both measure multiple disease-linked loci as well as screen for an euploidy, from a single cell. In some embodiments, the

## US 11,519,035 B2

147

PARENTAL SUPPORT<sup>TM</sup> method may be used to characterize one or more cells from embryos biopsied during an IVF cycle to determine the genetic condition of the one or more cells.

The PARENTAL SUPPORT<sup>TM</sup> method allows the cleaning of noisy genetic data. This may be done by inferring the correct genetic alleles in the target genome (embryo) using the genotype of related individuals (parents) as a reference. PARENTAL SUPPORT™ may be particularly relevant where only a small quantity of genetic material is available 10 (e.g. PGD) and where direct measurements of the genotypes are inherently noisy due to the limited amounts of genetic material. PARENTAL SUPPORT<sup>TM</sup> may be particularly relevant where only a small fraction of the genetic material available is from the target individual (e.g. NPD) and where 15 direct measurements of the genotypes are inherently noisy due to the contaminating DNA signal from another individual. The PARENTAL SUPPORT<sup>TM</sup> method is able to reconstruct highly accurate ordered diploid allele sequences on the embryo, together with copy number of chromosomes 20 segments, even though the conventional, unordered diploid measurements may be characterized by high rates of allele dropouts, drop-ins, variable amplification biases and other errors. The method may employ both an underlying genetic model and an underlying model of measurement error. The 25 genetic model may determine both allele probabilities at each SNP and crossover probabilities between SNPs. Allele probabilities may be modeled at each SNP based on data obtained from the parents and model crossover probabilities between SNPs based on data obtained from the HapMap 30 database, as developed by the International HapMap Project. Given the proper underlying genetic model and measurement error model, maximum a posteriori (MAP) estimation may be used, with modifications for computationally efficiency, to estimate the correct, ordered allele values at each 3. SNP in the embryo.

The techniques outlined above, in some cases, are able to determine the genotype of an individual given a very small amount of DNA originating from that individual. This could be the DNA from one or a small number of cells, or it could 40 be from the small amount of fetal DNA found in maternal blood.

### Hypotheses

In the context of this disclosure, a hypothesis refers to a possible genetic state. It may refer to a possible ploidy state. 45 It may refer to a possible allelic state. A set of hypotheses may refer to a set of possible genetic states, a set of possible allelic states, a set of possible ploidy states, or combinations thereof. In some embodiments, a set of hypotheses may be designed such that one hypothesis from the set will correspond to the actual genetic state of any given individual. In some embodiments, a set of hypotheses may be designed such that every possible genetic state may be described by at least one hypothesis from the set. In some embodiments of the present disclosure, one aspect of a method is to 55 determine which hypothesis corresponds to the actual genetic state of the individual in question.

In another embodiment of the present disclosure, one step involves creating a hypothesis. In some embodiments it may be a copy number hypothesis. In some embodiments it may 60 involve a hypothesis concerning which segments of a chromosome from each of the related individuals correspond genetically to which segments, if any, of the other related individuals. Creating a hypothesis may refer to the act of setting the limits of the variables such that the entire set of 65 possible genetic states that are under consideration are encompassed by those variables.

148

A "copy number hypothesis," also called a "ploidy hypothesis," or a "ploidy state hypothesis," may refer to a hypothesis concerning a possible ploidy state for a given chromosome copy, chromosome type, or section of a chromosome, in the target individual. It may also refer to the ploidy state at more than one of the chromosome types in the individual. A set of copy number hypotheses may refer to a set of hypotheses where each hypothesis corresponds to a different possible ploidy state in an individual. A set of hypotheses may concern a set of possible ploidy states, a set of possible parental haplotypes contributions, a set of possible fetal DNA percentages in the mixed sample, or combinations thereof. In some embodiments, the copy number hypotheses include all fetuses in a multiple pregnancy being euploid, all fetuses in a multiple pregnancy being aneuploid (such as any of the aneuploidies disclosed herein), and/or one or more fetuses in a multiple pregnancy being euploid and one or more fetuses in a multiple pregnancy being aneuploidy. In some embodiments, the copy number hypotheses include identical twins (also referred to as monozygotic twins) or fraternal twins (also referred to as dizygotic twins). In some embodiments, the copy number hypotheses include a molar pregnancy, such as a complete or partial molar

A normal individual contains one of each chromosome type from each parent. However, due to errors in meiosis and mitosis, it is possible for an individual to have 0, 1, 2, or more of a given chromosome type from each parent. In practice, it is rare to see more that two of a given chromosomes from a parent. In this disclosure, some embodiments only consider the possible hypotheses where 0, 1, or 2 copies of a given chromosome come from a parent; it is a trivial extension to consider more or less possible copies originating from a parent. In some embodiments, for a given chromosome, there are nine possible hypotheses: the three possible hypothesis concerning 0, 1, or 2 chromosomes of maternal origin, multiplied by the three possible hypotheses concerning 0, 1, or 2 chromosomes of paternal origin. Let (m,f) refer to the hypothesis where m is the number of a given chromosome inherited from the mother, and f is the number of a given chromosome inherited from the father. Therefore, the nine hypotheses are (0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), and (2,2). These may also be written as  $H_{00}$ ,  $H_{01}$ ,  $H_{02}$ ,  $H_{10}$ ,  $H_{12}$ ,  $H_{20}$ ,  $H_{21}$ , and  $H_{22}$ . The different hypotheses correspond to different ploidy states. For example, (1,1) refers to a normal disomic chromosome; (2,1) refers to a maternal trisomy, and (0,1) refers to a paternal monosomy. In some embodiments, the case where two chromosomes are inherited from one parent and one chromosome is inherited from the other parent may be further differentiated into two cases: one where the two chromosomes are identical (matched copy error), and one where the two chromosomes are homologous but not identical (unmatched copy error). In these embodiments, there are sixteen possible hypotheses. It should be understood that it is possible to use other sets of hypotheses, and a different number of hypotheses.

In some embodiments of the present disclosure, the ploidy hypothesis refers to a hypothesis concerning which chromosome from other related individuals correspond to a chromosome found in the target individual's genome. In some embodiments, a key to the method is the fact that related individuals can be expected to share haplotype blocks, and using measured genetic data from related individuals, along with a knowledge of which haplotype blocks match between the target individual and the related individual, it is possible to infer the correct genetic data for a

## US 11,519,035 B2

149

target individual with higher confidence than using the target individual's genetic measurements alone. As such, in some embodiments, the ploidy hypothesis may concern not only the number of chromosomes, but also which chromosomes in related individuals are identical, or nearly identical, with one or more chromosomes in the target individual.

Once the set of hypotheses have been defined, when the algorithms operate on the input genetic data, they may output a determined statistical probability for each of the hypotheses under consideration. The probabilities of the various hypotheses may be determined by mathematically calculating, for each of the various hypotheses, the value that the probability equals, as stated by one or more of the expert techniques, algorithms, and/or methods described elsewhere in this disclosure, using the relevant genetic data as input.

Once the probabilities of the different hypotheses are estimated, as determined by a plurality of techniques, they may be combined. This may entail, for each hypothesis, 20 multiplying the probabilities as determined by each technique. The product of the probabilities of the hypotheses may be normalized. Note that one ploidy hypothesis refers to one possible ploidy state for a chromosome.

The process of "combining probabilities," also called 25 "combining hypotheses," or combining the results of expert techniques, is a concept that should be familiar to one skilled in the art of linear algebra. One possible way to combine probabilities is as follows: When an expert technique is used to evaluate a set of hypotheses given a set of genetic data, 30 the output of the method is a set of probabilities that are associated, in a one-to-one fashion, with each hypothesis in the set of hypotheses. When a set of probabilities that were determined by a first expert technique, each of which are associated with one of the hypotheses in the set, are combined with a set of probabilities that were determined by a second expert technique, each of which are associated with the same set of hypotheses, then the two sets of probabilities are multiplied. This means that, for each hypothesis in the set, the two probabilities that are associated with that 40 hypothesis, as determined by the two expert methods, are multiplied together, and the corresponding product is the output probability. This process may be expanded to any number of expert techniques. If only one expert technique is used, then the output probabilities are the same as the input 45 probabilities. If more than two expert techniques are used, then the relevant probabilities may be multiplied at the same time. The products may be normalized so that the probabilities of the hypotheses in the set of hypotheses sum to 100%.

In some embodiments, if the combined probabilities for a 50 given hypothesis are greater than the combined probabilities for any of the other hypotheses, then it may be considered that that hypothesis is determined to be the most likely. In some embodiments, a hypothesis may be determined to be the most likely, and the ploidy state, or other genetic state, 55 may be called if the normalized probability is greater than a threshold. In an embodiment, this may mean that the number and identity of the chromosomes that are associated with that hypothesis may be called as the ploidy state. In an embodiment, this may mean that the identity of the alleles 60 that are associated with that hypothesis may be called as the allelic state. In some embodiments, the threshold may be between about 50% and about 80%. In some embodiments the threshold may be between about 80% and about 90%. In some embodiments the threshold may be between about 65 90% and about 95%. In some embodiments the threshold may be between about 95% and about 99%. In some

150

embodiments the threshold may be between about 99% and about 99.9%. In some embodiments the threshold may be above about 99.9%.

Parental Contexts

The parental context refers to the genetic state of a given allele, on each of the two relevant chromosomes for one or both of the two parents of the target. Note that in an embodiment, the parental context does not refer to the allelic state of the target, rather, it refers to the allelic state of the parents. The parental context for a given SNP may consist of four base pairs, two paternal and two maternal; they may be the same or different from one another. It is typically written as " $m_1m_2f_1f_2$ ," where  $m_1$  and  $m_2$  are the genetic state of the given SNP on the two maternal chromosomes, and f1 and f2 are the genetic state of the given SNP on the two paternal chromosomes. In some embodiments, the parental context may be written as " $f_1f_2|m_1m_2$ ." Note that subscripts "1" and "2" refer to the genotype, at the given allele, of the first and second chromosome; also note that the choice of which chromosome is labeled "1" and which is labeled "2" is arbitrary.

Note that in this disclosure, A and B are often used to generically represent base pair identities; A or B could equally well represent C (cytosine), G (guanine), A (adenine) or T (thymine). For example, if, at a given SNP based allele, the mother's genotype was T at that SNP on one chromosome, and G at that SNP on the homologous chromosome, and the father's genotype at that allele is G at that SNP on both of the homologous chromosomes, one may say that the target individual's allele has the parental context of AB|BB; it could also be said that the allele has the parental context of AB|AA. Note that, in theory, any of the four possible nucleotides could occur at a given allele, and thus it is possible, for example, for the mother to have a genotype of AT, and the father to have a genotype of GC at a given allele. However, empirical data indicate that in most cases only two of the four possible base pairs are observed at a given allele. It is possible, for example when using single tandem repeats, to have more than two parental, more than four and even more than ten contexts. In this disclosure the discussion assumes that only two possible base pairs will be observed at a given allele, although the embodiments disclosed herein could be modified to take into account the cases where this assumption does not hold.

A "parental context" may refer to a set or subset of target SNPs that have the same parental context. For example, if one were to measure 1000 alleles on a given chromosome on a target individual, then the context AA|BB could refer to the set of all alleles in the group of 1,000 alleles where the genotype of the mother of the target was homozygous, and the genotype of the father of the target is homozygous, but where the maternal genotype and the paternal genotype are dissimilar at that locus. If the parental data is not phased, and thus AB=BA, then there are nine possible parental contexts: AAIAA, AAIAB, AAIBB, ABIAA, ABIAB, ABIBB, BB|AA, BB|AB, and BB|BB. If the parental data is phased, and thus AB BA, then there are sixteen different possible parental contexts: AA|AA, AA|AB, AA|BA, AA|BB, AB|AA, AB|AB, AB|BA, AB|BB, BA|AA, BA|AB, BAIBA, BAIBB, BBIAA, BBIAB, BBIBA, and BBIBB. Every SNP allele on a chromosome, excluding some SNPs on the sex chromosomes, has one of these parental contexts. The set of SNPs wherein the parental context for one parent is heterozygous may be referred to as the heterozygous

## US 11,519,035 B2

151

Use of Parental Contexts in NPD

Non-invasive prenatal diagnosis is an important technique that can be used to determine the genetic state of a fetus from genetic material that is obtained in a non-invasive manner, for example from a blood draw on the pregnant mother. The blood could be separated and the plasma isolated, followed by isolation of the plasma DNA. Size selection could be used to isolate the DNA of the appropriate length. The DNA may be preferentially enriched at a set of loci. This DNA can then be measured by a number of means, such as by hybridizing to a genotyping array and measuring the fluorescence, or by sequencing on a high throughput sequencer.

When sequencing is used for ploidy calling of a fetus in the context of non-invasive prenatal diagnosis, there are a number of ways to use the sequence data. The most common way one could use the sequence data is to simply count the number of reads that map to a given chromosome. For example, imagine if you are trying to determine the ploidy state of chromosome 21 on the fetus. Further imagine that 20 the DNA in the sample is comprised of 10% DNA of fetal origin, and 90% DNA of maternal origin. In this case, you could look at the average number of reads on a chromosome which can be expected to be disomic, for example chromosome 3, and compare that to the number of read on chro- 25 mosome 21, where the reads are adjusted for the number of base pairs on that chromosome that are part of a unique sequence. If the fetus were euploid, one would expect the amount of DNA per unit of genome to be about equal at all locations (subject to stochastic variations). On the other hand, if the fetus were trisomic at chromosome 21, then one would expect there to be more slightly more DNA per genetic unit from chromosome 21 than the other locations on the genome. Specifically one would expect there to be about 5% more DNA from chromosome 21 in the mixture. When sequencing is used to measure the DNA, one would expect about 5% more uniquely mappable reads from chromosome 21 per unique segment than from the other chromosomes. One could use the observation of an amount of DNA from 40 a particular chromosome that is higher than a certain threshold, when adjusted for the number of sequences that are uniquely mappable to that chromosome, as the basis for an aneuploidy diagnosis. Another method that may be used to detect aneuploidy is similar to that above, except that 45 parental contexts could be taken into account.

When considering which alleles to target, one may consider the likelihood that some parental contexts are likely to be more informative than others. For example, AAIBB and the symmetric context BB|AA are the most informative 50 contexts, because the fetus is known to carry an allele that is different from the mother. For reasons of symmetry, both AA|BB and BB|AA contexts may be referred to as AA|BB. Another set of informative parental contexts are AA|AB and BB|AB, because in these cases the fetus has a 50% chance 55 of carrying an allele that the mother does not have. For reasons of symmetry, both AA|AB and BB|AB contexts may be referred to as AA|AB. A third set of informative parental contexts are AB|AA and AB|BB, because in these cases the fetus is carrying a known paternal allele, and that allele is 60 also present in the maternal genome. For reasons of symmetry, both ABIAA and ABIBB contexts may be referred to as AB|AA. A fourth parental context is AB|AB where the fetus has an unknown allelic state, and whatever the allelic state, it is one in which the mother has the same alleles. The 65 fifth parental context is AA|AA, where the mother and father are heterozygous.

152

Different Implementations of the Presently Disclosed Embodiments

Methods are disclosed herein for determining the ploidy state of a target individual. The target individual may be a blastomere, an embryo, or a fetus. In some embodiments of the present disclosure, a method for determining the ploidy state of one or more chromosome in a target individual may include any of the steps described in this document, and combinations thereof:

In some embodiments the source of the genetic material to be used in determining the genetic state of the fetus may be fetal cells, such as nucleated fetal red blood cells, isolated from the maternal blood. The method may involve obtaining a blood sample from the pregnant mother. The method may involve isolating a fetal red blood cell using visual techniques, based on the idea that a certain combination of colors are uniquely associated with nucleated red blood cell, and a similar combination of colors is not associated with any other present cell in the maternal blood. The combination of colors associated with the nucleated red blood cells may include the red color of the hemoglobin around the nucleus, which color may be made more distinct by staining, and the color of the nuclear material which can be stained, for example, blue. By isolating the cells from maternal blood and spreading them over a slide, and then identifying those points at which one sees both red (from the Hemoglobin) and blue (from the nuclear material) one may be able to identify the location of nucleated red blood cells. One may then extract those nucleated red blood cells using a micromanipulator, use genotyping and/or sequencing techniques to measure aspects of the genotype of the genetic material in those cells.

In an embodiment, one may stain the nucleated red blood cell with a die that only fluoresces in the presence of fetal hemoglobin and not maternal hemoglobin, and so remove the ambiguity between whether a nucleated red blood cell is derived from the mother or the fetus. Some embodiments of the present disclosure may involve staining or otherwise marking nuclear material. Some embodiments of the present disclosure may involve specifically marking fetal nuclear material using fetal cell specific antibodies.

There are many other ways to isolate fetal cells from maternal blood, or fetal DNA from maternal blood, or to enrich samples of fetal genetic material in the presence of maternal genetic material. Some of these methods are listed here, but this is not intended to be an exhaustive list. Some appropriate techniques are listed here for convenience: using fluorescently or otherwise tagged antibodies, size exclusion chromatography, magnetically or otherwise labeled affinity tags, epigenetic differences, such as differential methylation between the maternal and fetal cells at specific alleles, density gradient centrifugation succeeded by CD45/14 depletion and CD71-positive selection from CD45/14 negative-cells, single or double Percoll gradients with different osmolalities, or galactose specific lectin method.

In an embodiment of the present disclosure, the target individual is a fetus, and the different genotype measurements are made on a plurality of DNA samples from the fetus. In some embodiments of the present disclosure, the fetal DNA samples are from isolated fetal cells where the fetal cells may be mixed with maternal cells. In some embodiments of the present disclosure, the fetal DNA samples are from free floating fetal DNA, where the fetal DNA may be mixed with free floating maternal DNA. In some embodiments, the fetal DNA samples may be derived from maternal plasma or maternal blood that contains a mixture of maternal DNA and fetal DNA. In some embodiments, the fetal DNA may be mixed with maternal DNA in maternal:fetal ratios ranging from 99.9:0.1% to 99:1%;

## US 11,519,035 B2

15

99:1% to 90:10%; 90:10% to 80:20%; 80:20% to 70:30%; 70:30% to 50:50%; 50:50% to 10:90%; or 10:90% to 1:99%; 1:99% to 0.1:99.9%.

The genetic data of the target individual and/or of the related individual can be transformed from a molecular state 5 to an electronic state by measuring the appropriate genetic material using tools and or techniques taken from a group including, but not limited to: genotyping microarrays, and high throughput sequencing. Some high throughput sequencing methods include Sanger DNA sequencing, pyrosequencing, the ILLUMINA SOLEXA platform, ILLUMINA'S GENOME ANALYZER, or APPLIED BIOSYSTEM'S 454 sequencing platform, HELICOS'S TRUE SINGLE MOLECULE SEQUENCING platform, HALCYON MOLECULAR'S electron microscope sequencing 15 method, or any other sequencing method.

All of these methods physically transform the genetic data stored in a sample of DNA into a set of genetic data that is typically stored in a memory device in route to being processed.

A relevant individual's genetic data may be measured by analyzing substances taken from a group including, but not limited to: the individual's bulk diploid tissue, one or more diploid cells from the individual, one or more haploid cells from the individual, one or more blastomeres from the target 25 individual, extra-cellular genetic material found on the individual, extra-cellular genetic material from the individual found in maternal blood, cells from the individual found in maternal blood, one or more embryos created from (a) gamete(s) from the related individual, one or more blastomeres taken from such an embryo, extra-cellular genetic material found on the related individual, genetic material known to have originated from the related individual, and combinations thereof.

In some embodiments, a set of at least one ploidy state 35 hypothesis may be created for each of the chromosomes types of interest of the target individual. Each of the ploidy state hypotheses may refer to one possible ploidy state of the chromosome or chromosome segment of the target individual. The set of hypotheses may include some or all of the 40 possible ploidy states that the chromosome of the target individual may be expected to have. Some of the possible ploidy states may include nullsomy, monosomy, disomy, uniparental disomy, euploidy, trisomy, matching trisomy, unmatching trisomy, maternal trisomy, paternal trisomy, 4. tetrasomy, balanced (2:2) tetrasomy, unbalanced (3:1) tetrasomy, pentasomy, hexasomy, other aneuploidy, and combinations thereof. Any of these aneuploidy states may be mixed or partial aneuploidy such as unbalanced translocations, balanced translocations, Robertsonian translocations, 50 recombinations, deletions, insertions, crossovers, and combinations thereof.

In some embodiments, the knowledge of the determined ploidy state may be used to make a clinical decision. This knowledge, typically stored as a physical arrangement of 55 matter in a memory device, may then be transformed into a report. The report may then be acted upon. For example, the clinical decision may be to terminate the pregnancy; alternately, the clinical decision may be to continue the pregnancy. In some embodiments the clinical decision may 60 involve an intervention designed to decrease the severity of the phenotypic presentation of a genetic disorder, or a decision to take relevant steps to prepare for a special needs child.

In an embodiment of the present disclosure, any of the 65 methods described herein may be modified to allow for multiple targets to come from same target individual, for

154

example, multiple blood draws from the same pregnant mother. This may improve the accuracy of the model, as multiple genetic measurements may provide more data with which the target genotype may be determined. In an embodiment, one set of target genetic data served as the primary data which was reported, and the other served as data to double-check the primary target genetic data. In an embodiment, a plurality of sets of genetic data, each measured from genetic material taken from the target individual, are considered in parallel, and thus both sets of target genetic data serve to help determine which sections of parental genetic data, measured with high accuracy, composes the fetal genome.

In an embodiment, the method may be used for the purpose of paternity testing. For example, given the SNP-based genotypic information from the mother, and from a man who may or may not be the genetic father, and the measured genotypic information from the mixed sample, it is possible to determine if the genotypic information of the male indeed represents that actual genetic father of the gestating fetus. A simple way to do this is to simply look at the contexts where the mother is AA, and the possible father is AB or BB. In these cases, one may expect to see the father contribution half (AA|AB) or all (AA|BB) of the time, respectively. Taking into account the expected ADO, it is straightforward to determine whether or not the fetal SNPs that are observed are correlated with those of the possible father.

One embodiment of the present disclosure could be as follows: a pregnant woman wants to know if her fetus is afflicted with Down Syndrome, and/or if it will suffer from Cystic Fibrosis, and she does not wish to bear a child that is afflicted with either of these conditions. A doctor takes her blood, and stains the hemoglobin with one marker so that it appears clearly red, and stains nuclear material with another marker so that it appears clearly blue. Knowing that maternal red blood cells are typically anuclear, while a high proportion of fetal cells contain a nucleus, the doctor is able to visually isolate a number of nucleated red blood cells by identifying those cells that show both a red and blue color. The doctor picks up these cells off the slide with a micromanipulator and sends them to a lab which amplifies and genotypes ten individual cells. By using the genetic measurements, the PARENTAL SUPPORT<sup>TM</sup> method is able to determine that six of the ten cells are maternal blood cells, and four of the ten cells are fetal cells. If a child has already been born to a pregnant mother, PARENTAL SUPPORT™ can also be used to determine that the fetal cells are distinct from the cells of the born child by making reliable allele calls on the fetal cells and showing that they are dissimilar to those of the born child. Note that this method is similar in concept to the paternal testing embodiment of the present disclosure. The genetic data measured from the fetal cells may be of very poor quality, comprising many allele drop outs, due to the difficulty of genotyping single cells. The clinician is able to use the measured fetal DNA along with the reliable DNA measurements of the parents to infer aspects of the genome of the fetus with high accuracy using PARENTAL SUPPORT<sup>TM</sup>, thereby transforming the genetic data contained on genetic material from the fetus into the predicted genetic state of the fetus, stored on a computer. The clinician is able to determine both the ploidy state of the fetus, and the presence or absence of a plurality of diseaselinked genes of interest. It turns out that the fetus is euploid, and is not a carrier for cystic fibrosis, and the mother decides to continue the pregnancy.

## US 11,519,035 B2

155

In an embodiment of the present disclosure, a pregnant mother would like to determine if her fetus is afflicted with any whole chromosomal abnormalities. She goes to her doctor, and gives a sample of her blood, and she and her husband gives samples of their own DNA from cheek swabs. 5 A laboratory researcher genotypes the parental DNA using the MDA protocol to amplify the parental DNA, and ILLU-MINA INFINIUM arrays to measure the genetic data of the parents at a large number of SNPs. The researcher then spins down the blood, takes the plasma, and isolates a sample of 10 free-floating DNA using size exclusion chromatography. Alternately, the researcher uses one or more fluorescent antibodies, such as one that is specific to fetal hemoglobin to isolate a nucleated fetal red blood cell. The researcher then takes the isolated or enriched fetal genetic material and 15 amplifies it using a library of 70-mer oligonucleotides appropriately designed such that two ends of each oligonucleotide corresponded to the flanking sequences on either side of a target allele. Upon addition of a polymerase, ligase, and the appropriate reagents, the oligonucleotides under- 20 went gap-filling circularization, capturing the desired allele. An exonuclease was added, heat-inactivated, and the products were used directly as a template for PCR amplification. The PCR products were sequenced on an ILLUMINA GENOME ANALYZER. The sequence reads were used as 25 input for the PARENTAL SUPPORT™ method, which then predicted the ploidy state of the fetus.

In another embodiment, a couple—where the mother, who is pregnant, and is of advanced maternal age-wants to know whether the gestating fetus has Down syndrome, 30 Turner Syndrome, Prader Willi syndrome, or some other whole chromosomal abnormality. The obstetrician takes a blood draw from the mother and father. The blood is sent to a laboratory, where a technician centrifuges the maternal sample to isolate the plasma and the buffy coat. The DNA in 3: the buffy coat and the paternal blood sample are transformed through amplification and the genetic data encoded in the amplified genetic material is further transformed from molecularly stored genetic data into electronically stored genetic data by running the genetic material on a high 40 throughput sequencer to measure the parental genotypes. The plasma sample is preferentially enriched at a set of loci using a 5,000-plex hemi-nested targeted PCR method. The mixture of DNA fragments is prepared into a DNA library suitable for sequencing. The DNA is then sequenced using 45 a high throughput sequencing method, for example, the ILLUMINA GAIIx GENOME ANALYZER. The sequencing transforms the information that is encoded molecularly in the DNA into information that is encoded electronically in computer hardware. An informatics based technique that 50 includes the presently disclosed embodiments, such as PARENTAL SUPPORT<sup>TM</sup>, may be used to determine the ploidy state of the fetus. This may involve calculating, on a computer, allele count probabilities at the plurality of polymorphic loci from the DNA measurements made on the 55 prepared sample; creating, on a computer, a plurality of ploidy hypotheses each pertaining to a different possible ploidy state of the chromosome; building, on a computer, a joint distribution model for the expected allele counts at the plurality of polymorphic loci on the chromosome for each 60 ploidy hypothesis; determining, on a computer, a relative probability of each of the ploidy hypotheses using the joint distribution model and the allele counts measured on the prepared sample; and calling the ploidy state of the fetus by selecting the ploidy state corresponding to the hypothesis 65 with the greatest probability. It is determined that the fetus has Down syndrome. A report is printed out, or sent elec156

tronically to the pregnant woman's obstetrician, who transmits the diagnosis to the woman. The woman, her husband, and the doctor sit down and discuss their options. The couple decides to terminate the pregnancy based on the knowledge that the fetus is afflicted with a trisomic condition.

In an embodiment, a company may decide to offer a diagnostic technology designed to detect aneuploidy in a gestating fetus from a maternal blood draw. Their product may involve a mother presenting to her obstetrician, who may draw her blood. The obstetrician may also collect a genetic sample from the father of the fetus. A clinician may isolate the plasma from the maternal blood, and purify the DNA from the plasma. A clinician may also isolate the buffy coat layer from the maternal blood, and prepare the DNA from the buffy coat. A clinician may also prepare the DNA from the paternal genetic sample. The clinician may use molecular biology techniques described in this disclosure to append universal amplification tags to the DNA in the DNA derived from the plasma sample. The clinician may amplify the universally tagged DNA. The clinician may preferentially enrich the DNA by a number of techniques including capture by hybridization and targeted PCR. The targeted PCR may involve nesting, hemi-nesting or semi-nesting, or any other approach to result in efficient enrichment of the plasma derived DNA. The targeted PCR may be massively multiplexed, for example with 10,000 primers in one reaction volume, where the primers target SNPs on chromosomes 13, 18, 21, X and those loci that are common to both X and Y, and optionally other chromosomes as well. The selective enrichment and/or amplification may involve tagging each individual molecule with different tags, molecular barcodes, tags for amplification, and/or tags for sequencing. The clinician may then sequence the plasma sample, and also possibly also the prepared maternal and/or paternal DNA. The molecular biology steps may be executed either wholly or partly by a diagnostic box. The sequence data may be fed into a single computer, or to another type of computing platform such as may be found in 'the cloud'. The computing platform may calculate allele counts at the targeted polymorphic loci from the measurements made by the sequencer. The computing platform may create a plurality of ploidy hypotheses pertaining to nullsomy, monosomy, disomy, matched trisomy, and unmatched trisomy for each of chromosomes 13, 18, 21, X and Y. The computing platform may build a joint distribution model for the expected allele counts at the targeted loci on the chromosome for each ploidy hypothesis for each of the five chromosomes being interrogated. The computing platform may determine a probability that each of the ploidy hypotheses is true using the joint distribution model and the allele counts measured on the preferentially enriched DNA derived from the plasma sample. The computing platform may call the ploidy state of the fetus, for each of chromosome 13, 18, 21, X and Y by selecting the ploidy state corresponding to the germane hypothesis with the greatest probability. A report may be generated comprising the called ploidy states, and it may be sent to the obstetrician electronically, displayed on an output device, or a printed hard copy of the report may be delivered to the obstetrician. The obstetrician may inform the patient and optionally the father of the fetus, and they may decide which clinical options are open to them, and which is most desirable.

In another embodiment, a pregnant woman, hereafter referred to as "the mother" may decide that she wants to know whether or not her fetus(es) are carrying any genetic abnormalities or other conditions. She may want to ensure that there are not any gross abnormalities before she is

# US 11,519,035 B2

confident to continue the pregnancy. She may go to her obstetrician, who may take a sample of her blood. He may also take a genetic sample, such as a buccal swab, from her cheek. He may also take a genetic sample from the father of the fetus, such as a buccal swab, a sperm sample, or a blood sample. He may send the samples to a clinician. The clinician may enrich the fraction of free floating fetal DNA in the maternal blood sample. The clinician may enrich the fraction of enucleated fetal blood cells in the maternal blood sample. The clinician may use various aspects of the methods described herein to determine genetic data of the fetus. That genetic data may include the ploidy state of the fetus, and/or the identity of one or a number of disease linked alleles in the fetus. A report may be generated summarizing the results of the prenatal diagnosis. The report may be 15 transmitted or mailed to the doctor, who may tell the mother the genetic state of the fetus. The mother may decide to discontinue the pregnancy based on the fact that the fetus has one or more chromosomal, or genetic abnormalities, or undesirable conditions. She may also decide to continue the 20 pregnancy based on the fact that the fetus does not have any gross chromosomal or genetic abnormalities, or any genetic conditions of interest

Another example may involve a pregnant woman who has been artificially inseminated by a sperm donor, and is 25 pregnant. She wants to minimize the risk that the fetus she is carrying has a genetic disease. She has blood drawn at a phlebotomist, and techniques described in this disclosure are used to isolate three nucleated fetal red blood cells, and a tissue sample is also collected from the mother and genetic 30 father. The genetic material from the fetus and from the mother and father are amplified as appropriate and genotyped using the ILLUMINA INFINIUM BEADARRAY, and the methods described herein clean and phase the parental and fetal genotype with high accuracy, as well as to make 35 ploidy calls for the fetus. The fetus is found to be euploid, and phenotypic susceptibilities are predicted from the reconstructed fetal genotype, and a report is generated and sent to the mother's physician so that they can decide what clinical decisions may be best.

In an embodiment, the raw genetic material of the mother and the father is transformed by way of amplification to an amount of DNA that is similar in sequence, but larger in quantity. Then, by way of a genotyping method, the genotypic data that is encoded by nucleic acids is transformed 45 into genetic measurements that may be stored physically and/or electronically on a memory device, such as those described above. The relevant algorithms that makeup the PARENTAL SUPPORT<sup>TM</sup> algorithm, relevant parts of which are discussed in detail herein, are translated into a 50 computer program, using a programming language. Then, through the execution of the computer program on the computer hardware, instead of being physically encoded bits and bytes, arranged in a pattern that represents raw measurement data, they become transformed into a pattern that 55 represents a high confidence determination of the ploidy state of the fetus. The details of this transformation will rely on the data itself and the computer language and hardware system used to execute the method described herein. Then, the data that is physically configured to represent a high 60 quality ploidy determination of the fetus is transformed into a report which may be sent to a health care practitioner. This transformation may be carried out using a printer or a computer display. The report may be a printed copy, on paper or other suitable medium, or else it may be electronic. 65 In the case of an electronic report, it may be transmitted, it may be physically stored on a memory device at a location

on the computer accessible by the health care practitioner; it also may be displayed on a screen so that it may be read. In the case of a screen display, the data may be transformed to a readable format by causing the physical transformation of pixels on the display device. The transformation may be accomplished by way of physically firing electrons at a phosphorescent screen, by way of altering an electric charge that physically changes the transparency of a specific set of pixels on a screen that may lie in front of a substrate that emits or absorbs photons. This transformation may be accomplished by way of changing the nanoscale orientation of the molecules in a liquid crystal, for example, from nematic to cholesteric or smectic phase, at a specific set of pixels. This transformation may be accomplished by way of an electric current causing photons to be emitted from a specific set of pixels made from a plurality of light emitting diodes arranged in a meaningful pattern. This transformation may be accomplished by any other way used to display information, such as a computer screen, or some other output device or way of transmitting information. The health care practitioner may then act on the report, such that the data in the report is transformed into an action. The action may be to continue or discontinue the pregnancy, in which case a gestating fetus with a genetic abnormality is transformed into non-living fetus. The transformations listed herein may be aggregated, such that, for example, one may transform the genetic material of a pregnant mother and the father, through a number of steps outlined in this disclosure, into a medical decision consisting of aborting a fetus with genetic abnormalities, or consisting of continuing the pregnancy. Alternately, one may transform a set of genotypic measurements into a report that helps a physician treat his pregnant patient.

In an embodiment of the present disclosure, the method described herein can be used to determine the ploidy state of a fetus even when the host mother, i.e. the woman who is pregnant, is not the biological mother of the fetus she is carrying. In an embodiment of the present disclosure, the method described herein can be used to determine the ploidy state of a fetus using only the maternal blood sample, and without the need for a paternal genetic sample.

Some of the math in the presently disclosed embodiments makes hypotheses concerning a limited number of states of aneuploidy. In some cases, for example, only zero, one or two chromosomes are expected to originate from each parent. In some embodiments of the present disclosure, the mathematical derivations can be expanded to take into account other forms of aneuploidy, such as quadrosomy, where three chromosomes originate from one parent, pentasomy, hexasomy etc., without changing the fundamental concepts of the present disclosure. At the same time, it is possible to focus on a smaller number of ploidy states, for example, only trisomy and disomy. Note that ploidy determinations that indicate a non-whole number of chromosomes may indicate mosaicism in a sample of genetic material.

In some embodiments, the genetic abnormality is a type of aneuploidy, such as Down syndrome (or trisomy 21), Edwards syndrome (trisomy 18), Patau syndrome (trisomy 13), Turner Syndrome (45X), Klinefelter's syndrome (a male with 2 X chromosomes), Prader-Willi syndrome, and DiGeorge syndrome (UPD 15). Congenital disorders, such as those listed in the prior sentence, are commonly undesirable, and the knowledge that a fetus is afflicted with one or more phenotypic abnormalities may provide the basis for a decision to terminate the pregnancy, to take necessary precautions to prepare for the birth of a special needs child,

## US 11,519,035 B2

159

or to take some therapeutic approach meant to lessen the severity of a chromosomal abnormality.

In some embodiments, the methods described herein can be used at a very early gestational age, for example as early as four week, as early as five weeks, as early as six weeks, as early as seven weeks, as early as eight weeks, as early as nine weeks, as early as ten weeks, as early as eleven weeks, and as early as twelve weeks.

In some embodiments, a method disclosed herein is used in the context of pre-implantation genetic diagnosis (PGD) 10 for embryo selection during in vitro fertilization, where the target individual is an embryo, and the parental genotypic data can be used to make ploidy determinations about the embryo from sequencing data from a single or two cell biopsy from a day 3 embryo or a trophectoderm biopsy from 15 a day 5 or day 6 embryo. In a PGD setting, only the child DNA is measured, and only a small number of cells are tested, generally one to five but as many as ten, twenty or fifty. The total number of starting copies of the A and B alleles (at a SNP) are then trivially determined by the child 20 genotype and the number of cells. In NPD, the number of starting copies is very high and so the allele ratio after PCR is expected to accurately reflect the starting ratio. However, the small number of starting copies in PGD means that contamination and imperfect PCR efficiency have a non- 25 trivial effect on the allele ratio following PCR. This effect may be more important than depth of read in predicting the variance in the allele ratio measured after sequencing. The distribution of measured allele ratio given a known child genotype may be created by Monte Carlo simulation of the 30 PCR process based on the PCR probe efficiency and probability of contamination. Given an allele ratio distribution for each possible child genotype, the likelihoods of various hypotheses can be calculated as described for NIPD. Maximum Likelihood Estimates

Most methods known in the art for detecting the presence or absence of biological phenomenon or medical condition involve the use of a single hypothesis rejection test, where a metric that is correlated with the condition is measured, and if the metric is on one side of a given threshold, the condition is present, while of the metric falls on the other side of the threshold, the condition is absent. A single-hypothesis rejection test only looks at the null distribution when deciding between the null and alternate hypotheses. Without taking into account the alternate distribution, one 45 cannot estimate the likelihood of each hypothesis given the observed data and therefore cannot calculate a confidence on the call. Hence with a single-hypothesis rejection test, one gets a yes or no answer without a feeling for the confidence associated with the specific case.

In some embodiments, the method disclosed herein is able to detect the presence or absence of biological phenomenon or medical condition using a maximum likelihood method. This is a substantial improvement over a method using a single hypothesis rejection technique as the threshold for 55 calling absence or presence of the condition can be adjusted as appropriate for each case. This is particularly relevant for diagnostic techniques that aim to determine the presence or absence of aneuploidy in a gestating fetus from genetic data available from the mixture of fetal and maternal DNA 60 present in the free floating DNA found in maternal plasma. This is because as the fraction of fetal DNA in the plasma derived fraction changes, the optimal threshold for calling aneuploidy vs. euploidy changes. As the fetal fraction drops, the distribution of data that is associated with an aneuploidy becomes increasingly similar to the distribution of data that is associated with a euploidy.

160

The maximum likelihood estimation method uses the distributions associated with each hypothesis to estimate the likelihood of the data conditioned on each hypothesis. These conditional probabilities can then be converted to a hypothesis call and confidence. Similarly, maximum a posteriori estimation method uses the same conditional probabilities as the maximum likelihood estimate, but also incorporates population priors when choosing the best hypothesis and determining confidence.

Therefore, the use of a maximum likelihood estimate (MLE) technique, or the closely related maximum a posteriori (MAP) technique give two advantages, first it increases the chance of a correct call, and it also allows a confidence to be calculated for each call. In an embodiment, selecting the ploidy state corresponding to the hypothesis with the greatest probability is carried out using maximum likelihood estimates or maximum a posteriori estimates. In an embodiment, a method is disclosed for determining the ploidy state of a gestating fetus that involves taking any method currently known in the art that uses a single hypothesis rejection technique and reformulating it such that it uses a MLE or MAP technique. Some examples of methods that can be significantly improved by applying these techniques can be found in U.S. Pat. Nos. 8,008,018, 7,888,017, or U.S. Pat. No. 7,332,277

In an embodiment, a method is described for determining presence or absence of fetal aneuploidy in a maternal plasma sample comprising fetal and maternal genomic DNA, the method comprising: obtaining a maternal plasma sample; measuring the DNA fragments found in the plasma sample with a high throughput sequencer; mapping the sequences to the chromosome and determining the number of sequence reads that map to each chromosome; calculating the fraction of fetal DNA in the plasma sample; calculating an expected distribution of the amount of a target chromosome that would be expected to be present if that if the second target chromosome were euploid and one or a plurality of expected distributions that would be expected if that chromosome were aneuploid, using the fetal fraction and the number of sequence reads that map to one or a plurality of reference chromosomes expected to be euploid; and using a MLE or MAP determine which of the distributions is most likely to be correct, thereby indicating the presence or absence of a fetal aneuploidy. In an embodiment, the measuring the DNA from the plasma may involve conducting massively parallel shotgun sequencing. In an embodiment, the measuring the DNA from the plasma sample may involve sequencing DNA that has been preferentially enriched, for example through targeted amplification, at a plurality of polymorphic or non-polymorphic loci. The plurality of loci may be designed to target one or a small number of suspected aneuploid chromosomes and one or a small number of reference chromosomes. The purpose of the preferential enrichment is to increase the number of sequence reads that are informative for the ploidy determination.

Ploidy Calling Informatics Methods

Described herein is a method for determining the ploidy state of a fetus given sequence data. In some embodiments, this sequence data may be measured on a high throughput sequencer. In some embodiments, the sequence data may be measured on DNA that originated from free floating DNA isolated from maternal blood, wherein the free floating DNA comprises some DNA of maternal origin, and some DNA of fetal/placental origin. This section will describe one embodiment of the present disclosure in which the ploidy state of the fetus is determined assuming that fraction of fetal DNA in the mixture that has been analyzed is not known and will

16

be estimated from the data. It will also describe an embodiment in which the fraction of fetal DNA ("fetal fraction") or the percentage of fetal DNA in the mixture can be measured by another method, and is assumed to be known in determining the ploidy state of the fetus. In some embodiments 5 the fetal fraction can be calculated using only the genotyping measurements made on the maternal blood sample itself, which is a mixture of fetal and maternal DNA. In some embodiments the fraction may be calculated also using the measured or otherwise known genotype of the mother and/or the measured or otherwise known genotype of the father. In another embodiment ploidy state of the fetus can be determined solely based on the calculated fraction of fetal DNA for the chromosome in question compared to the calculated fraction of fetal DNA for the reference chromosome assumed disomic.

In the preferred embodiment, suppose that, for a particular chromosome, we observe and analyze N SNPs, for which we have:

Set of NR free floating DNA sequence measurements  $S=(s_1,\ldots,s_{NR})$ . Since this method utilizes the SNP measurements, all sequence data that corresponds to non-polymorphic loci can be disregarded. In a simplified version, where we have (A,B) counts on each SNP, where A and B correspond to the two alleles present at a given locus, S can be written as  $S=((a_1,b_1),\ldots,(a_N,b_N))$ , where  $a_i$  is the A count on SNP i,  $b_i$  is the B count on SNP i, and  $\sum_{i=1:N}(a_i+b_i)=NR$ 

Parent data consisting of

Genotypes from a SNP microarray or other intensity based genotyping platform: mother  $M=(m_1,\ldots,m_N)$ , father  $F=(f_1,\ldots,f_N)$ , where  $m_i$ ,  $f_i\in (AA,AB,BB)$ .

AND/OR sequence data measurements: NRM mother measurements SM= $(sm_1, \ldots, sm_{nrm})$ , NRF father measurements SF= $(sf_1, \ldots, sf_{nrf})$ . Similar to the above simplification, if we have (A,B) counts on each SNP SM= $((am_1, bm_1), \ldots, (am_N, bm_N))$ , SF= $((af_1, bf_1), \ldots, (af_N, bf_N))$ 

Collectively, the mother, father child data are denoted as D=(M,F,SM,SF,S). Note that the parent data is desired and increases the accuracy of the algorithm, but is NOT necessary, especially the father data. This means that even in the absence of mother and/or father data, it is possible to get very accurate copy number results.

It is possible to derive the best copy number estimate (H\*) by maximizing the data log likelihood LIK(DIH) over all hypotheses (H) considered. In particular it is possible to determine the relative probability of each of the ploidy hypotheses using the joint distribution model and the allele counts measured on the prepared sample, and using those relative probabilities to determine the hypothesis most likely to be correct as follows:

$$H^* = \underset{H}{\operatorname{argmax}} \ LIK(D \mid H)$$

Similarly the a posteriori hypothesis likelihood given the data may be written as:

$$H^* = \underset{\alpha}{\operatorname{argmax}} \ LIK(D \mid H) * priorprob(H)$$

Where priorprob(H) is the prior probability assigned to 65 each hypothesis H, based on model design and prior knowledge.

162

It is also possible to use priors to find the maximum a posteriori estimate:

$$H_{MA} = \underset{H}{\operatorname{argmax}} LIK(D \mid H)$$

In an embodiment, the copy number hypotheses that may be considered are:

Monosomy:

maternal H10 (one copy from mother)

paternal H01 (one copy from father)

Disomy: H11 (one copy each mother and father)

Simple trisomy, no crossovers considered:

Maternal: H21\_matched (two identical copies from mother, one copy from father),

H21\_unmatched (BOTH copies from mother, one copy from father)

Paternal: H12\_matched (one copy from mother, two identical copies from father),

H12\_unmatched (one copy from mother, both copies from father)

Composite trisomy, allowing for crossovers (using a joint distribution model):

maternal H21 (two copies from mother, one from father), paternal H12 (one copy from mother, two copies from father)

In other embodiments, other ploidy states, such as nullsomy (H00), uniparental disomy (H20 and H02), and tetrasomy (H04, H13, H22, H31 and H40), may be considered.

If there are no crossovers, each trisomy, whether the origin was mitosis, meiosis I, or meiosis II, would be one of the matched or unmatched trisomies. Due to crossovers, true trisomy is usually a combination of the two. First, a method to derive hypothesis likelihoods for simple hypotheses is described. Then a method to derive hypothesis likelihoods for composite hypotheses is described, combining individual SNP likelihood with crossovers.

LIK(D|H) for a Simple Hypothesis

In an embodiment, LIK(DIH) may be determined for simple hypotheses, as follows. For simple hypotheses H, LIK(H), the log likelihood of hypothesis H on a whole chromosome, may be calculated as the sum of log likelihoods of individual SNPs, assuming known or derived child fraction cf. In an embodiment it is possible to derive cf from the data.

$$LIK(D\mid H) = \sum_{i} LIK(D\mid H,\,cf,\,i)$$

This hypothesis does not assume any linkage between SNPs, and therefore does not utilize a joint distribution model.

In some embodiments, the Log Likelihood may be determined on a per SNP basis. On a particular SNP i, assuming fetal ploidy hypothesis H and percent fetal DNA cf, log likelihood of observed data D is defined as:

$$\begin{split} LIK(D\mid H,\,i) &= \log\,P(D\mid H,\,cf,\,i) = \\ &\log \left(\sum_{m,f,c}P(D\mid m,\,f,\,c,\,H,\,cf,\,i)P(c\mid m,\,f,\,H)P(m\mid i)P(f\mid i)\right) \end{split}$$

163

where m are possible true mother genotypes, f are possible true father genotypes, where  $m,f \in \{AA,AB,BB\}$ , and c are possible child genotypes given the hypothesis H. In particular, for monosomy  $c \in \{AA, B\}$ , for disomy  $c \in \{AA, AB, BB\}$ , for trisomy  $c \in \{AAA, AAB, ABB, BBB\}$ . Genotype prior frequency: p(m|i) is the general prior probability of mother genotype m on SNP i, based on the known population frequency at SNP I, denoted pA. In particular

$$\begin{array}{c} p(AA|pA_i) = (pA_i)^2, p(AB|pA_i) = 2(pA_i)^*(1-pA_i), p \\ (BB|pA_i) = (1-pA_i)^2 \end{array}$$

Father genotype probability, p(fli), may be determined in an analogous fashion.

True child probability: p(clm, f, H) is the probability of getting true child genotype=c, given parents m, f, and assuming hypothesis H, which can be easily calculated. For example, for H11, H21 matched and H21 unmatched, p(clm, f,H) is given below.

164

if no father genotype data F is available, one may just use  $P(F|f,\ i)=1$ . If no father sequence data SF is available, one may just use P(SF|f,i)=1.

In some embodiments, the method involves building a joint distribution model for the expected allele counts at a plurality of polymorphic loci on the chromosome for each ploidy hypothesis; one method to accomplish such an end is described here. Free fetal DNA data likelihood: P(S|m, c, H, cf, i) is the probability of free fetal DNA sequence data on SNP i, given true mother genotype m, true child genotype c, child copy number hypothesis H, and assuming child fraction cf. It is in fact the probability of sequence data S on SNP I, given the true probability of A content on SNP i μ(m, c, cf, H)

 $P(S|m,c,H,cf,i)=P(S|\mu(m,c,cf,H),i)$ 

	p(c m, f, H)											
		H11			H21 matched				H21 unmatched			
m	f	AA	AB	вв	AAA	AAB	ABB	BBB	AAA	AAB	ABB	BBB
AA	AA	1	0	0	1	0	0	0	1	0	0	0
AB	AA	0.5	0.5	0	0.5	0	0.5	0	0	1	0	0
$^{\mathrm{BB}}$	AA	0	1	0	0	0	1	0	0	0	1	0
AA	AB	0.5	0.5	0	0.5	0.5	0	0	0.5	0.5	0	0
AB	AB	0.25	0.5	0.25	0.25	0.25	0.25	0.25	0	0.5	0.5	0
BB	AB	0	0.5	0.5	0	0	0.5	0.5	0	0	0.5	0.5
AA	$^{\mathrm{BB}}$	0	1	0	0	1	0	0	0	1	0	0
AB	BB	0	0.5	0.5	0	0.5	0	0.5	0	0	1	0
BB	BB	0	0	1	0	0	0	1	0	0	0	1

Data likelihood: P(D|m, f, c, H, i, cf) is the probability of given data D on SNP i, given true mother genotype m, true father genotype f, true child genotype c, hypothesis H and child fraction cf. It can be broken down into the probability of mother, father and child data as follows:

$$P(D|\textit{m,f,c,H,cf,i}) = P(SM|\textit{m,i})P(M|\textit{m,i})P(SF|\textit{f,i})P(F|\textit{f,i}) \\ P(S|\textit{m,c,H,cf,i})$$

Mother SNP array data likelihood: Probability of mother SNP array genotype data  $m_i$  at SNP i compared to true genotype m, assuming SNP array genotypes are correct, is  $^{45}$  simply

$$P(M \mid m, i) = \begin{cases} 1 & m_i = m \\ 0 & m_i \neq m \end{cases}$$

Mother sequence data likelihood: the probability of the mother sequence data at SNP i, in the case of counts  $S_i$ =(am<sub>i</sub>,bm<sub>i</sub>), with no extra noise or bias involved, is the 55 binomial probability defined as  $P(SM|m,i)=P_{X|m}(am_i)$  where X|m-Binom( $p_m(A)$ , am<sub>i</sub>+bm<sub>i</sub>) with  $p_m(A)$  defined as

m	AA	AB	BB	В	В	nocall
p(A)	1	0.5	0	1	0	0.5

Father data likelihood: a similar equation applies for father data likelihood.

Note that it is possible to determine the child genotype without the parent data, especially father data. For example

For counts, where  $S_i=(a_i,b_i)$ , with no extra noise or bias in data involved,

$$P(S||\mu(m,c,cf,H),i) = P_x(a_i)$$

where X-Binom(p(A),  $a_i$ + $b_i$ ) with p(A)= $\mu$ (m, c, cf, H). In a more complex case where the exact alignment and (A,B) counts per SNP are not known, P(S| $\mu$ (m, c, cf, H),i) is a combination of integrated binomials.

True A content probability: p(m, c, cf, H), the true probability of A content on SNP i in this mother/child mixture, assuming that true mother genotype=m, true child genotype=c, and overall child fraction=cf, is defined as

$$\mu(m,\,c,\,cf\,,\,H) = \frac{\#A(m)*(1-cf) + \#A(c)*cf}{n_m*(1-cf) + n_c*cf}$$

where #A(g)=number of A's in genotype g,  $n_m$ =2 is somy of mother and n, is ploidy of the child under hypothesis H (1 for monosomy, 2 for disomy, 3 for trisomy).

Using a Joint Distribution Model: LIK(D $\mid$ H) for a Composite Hypothesis

In some embodiments, the method involves building a joint distribution model for the expected allele counts at the plurality of polymorphic loci on the chromosome for each ploidy hypothesis; one method to accomplish such an end is described here. In many cases, trisomy is usually not purely matched or unmatched, due to crossovers, so in this section results for composite hypotheses H21 (maternal trisomy) and H12 (paternal trisomy) are derived, which combine matched and unmatched trisomy, accounting for possible crossovers.

In the case of trisomy, if there were no crossovers, trisomy would be simply matched or unmatched trisomy. Matched trisomy is where child inherits two copies of the identical

165

chromosome segment from one parent. Unmatched trisomy is where child inherits one copy of each homologous chromosome segment from the parent. Due to crossovers, some segments of a chromosome may have matched trisomy, and other parts may have unmatched trisomy. Described in this section is how to build a joint distribution model for the heterozygosity rates for a set of alleles; that is, for the expected allele counts at a number of loci for one or more hypotheses.

Suppose that on SNP i, LIK(D|Hm,i) is the fit for matched 10 hypothesis  $H_m$ , and LIK(D|Hu, i) is the fit for unmatched hypothesis  $H_{uv}$  and pc(i) probability of crossover between SNPs i–1 and i. One may then calculate the full likelihood as:

$$LIK(D\mid H) = \sum_{E} LIK(D\mid E,\, 1:N)$$

where LIK(DIE, 1: N) is the likelihood of ending in hypothesis E, for SNPs 1:N. E=hypothesis of the last SNP, E $\in$ (Hm, Hu). Recursively, one may calculate:

$$\begin{split} LIK(D\mid E,\,1:i) &= LIK(D\mid E,\,i) + \log(\exp(LIK(D\mid E,\,1:i-1))*(1-pc(i)) + \\ &\quad \exp(LIK(D\mid \sim E,\,1:i-1))*pc(i)) \end{split}$$

where  $\sim$ E is the hypothesis other than E (not E), where 30 hypotheses considered are  $H_m$  and  $H_u$ . In particular, one may calculate the likelihood of 1: i SNPs, based on likelihood of 1 to (i–1) SNPs with either the same hypothesis and no crossover, or the opposite hypothesis and a crossover, multiplied by the likelihood of the SNP i 35

For SNP 1, i=1, LIK(DIE, 1:1)=LIK(DIE, 1). For SNP 2, i=2, LIK(DIE, 1:2)=LIK(DIE, 2)+log (exp(LIK (DIE, 1))\*(1-pc(2))+exp (LIK(DI $\sim$ E, 1))\*pc(2)), and so on for i=3·N.

In some embodiments, the child fraction may be determined. The child fraction may refer to the proportion of sequences in a mixture of DNA that originate from the child. In the context of non-invasive prenatal diagnosis, the child fraction may refer to the proportion of sequences in the maternal plasma that originate from the fetus or the portion of the placenta with fetal genotype. It may refer to the child fraction in a sample of DNA that has been prepared from the maternal plasma, and may be enriched in fetal DNA. One purpose of determining the child fraction in a sample of DNA is for use in an algorithm that can make ploidy calls on the fetus, therefore, the child fraction could refer to whatever sample of DNA was analyzed by sequencing for the purpose of non-invasive prenatal diagnosis.

Some of the algorithms presented in this disclosure that are part of a method of non-invasive prenatal aneuploidy diagnosis assume a known child fraction, which may not always the case. In an embodiment, it is possible to find the most likely child fraction by maximizing the likelihood for disomy on selected chromosomes, with or without the presence of the parental data

In particular, suppose that LIK(DIH11, cf, chr)=log likelihood as described above, for the disomy hypothesis, and for child fraction cf on chromosome chr. For selected chromosomes in Cset (usually 1:16), assumed to be euploid, the full likelihood is:

 $LIK(cf) = \sum_{chr \in Cset} Lik(D|H11,cf,chr)$ 

166

The most likely child fraction (cf\*) is derived as

$$cf^* = \underset{cf}{\operatorname{argmax}} LIK(cf).$$

It is possible to use any set of chromosomes. It is also possible to derive child fraction without assuming euploidy on the reference chromosomes. Using this method it is possible to determine the child fraction for any of the following situations: (1) one has array data on the parents and shotgun sequencing data on the maternal plasma; (2) one has array data on the parents and targeted sequencing data on the maternal plasma; (3) one has targeted sequencing data on both the parents and maternal plasma; (4) one has targeted sequencing data on both the mother and the maternal plasma fraction; (5) one has targeted sequencing data on the maternal plasma fraction; (6) other combinations of parental and child fraction measurements.

In some embodiments the informatics method may incorporate data dropouts; this may result in ploidy determinations of higher accuracy. Elsewhere in this disclosure it has been assumed that the probability of getting an A is a direct function of the true mother genotype, the true child genotype, the fraction of the child in the mixture, and the child copy number. It is also possible that mother or child alleles can drop out, for example instead of measuring true child AB in the mixture, it may be the case that only sequences mapping to allele A are measured. One may denote the parent dropout rate for genomic Illumina data dpg, parent dropout rate for sequence data  $d_{ns}$  and child dropout rate for sequence data  $d_{cs}$ . In some embodiments, the mother dropout rate may be assumed to be zero, and child dropout rates are relatively low; in this case, the results are not severely affected by dropouts. In some embodiments the possibility of allele dropouts may be sufficiently large that they result in a significant effect of the predicted ploidy call. For such a case, allele dropouts have been incorporated into the algorithm here:

Parent SNP array data dropouts: For mother genomic data M, suppose that the genotype after the dropout is  $m_d$ , then

$$P(M\mid m,\,i) = \sum_{m_d} P(M\mid m_d,\,i) P(m_d\mid m)$$

where

$$P(M \mid m_d, i) = \begin{cases} 1 & m_i = m_d \\ 0 & m_i \neq m_d \end{cases}$$

as before, and  $P(m_d|m)$  is the likelihood of genotype  $m_d$  after the possible dropout given the true genotype m, defined as below, for dropout rate d

00					md		
	m	AA	AB	ВВ	A	В	nocall
•	AA	(1 -	0	0	2d(1 -	0	d^2
55	AB	d)^2 0	(1 - d)^2	0	d) d(1 - d)	d(1 - d)	d^2

167 -continued

	md							
m	AA	AB	ВВ	A	В	nocall		
ВВ	0	0	(1 - d)^2	0	2d(1 - d)	d^2		

A similar equation applies for father SNP array data. Parent sequence data dropouts: For mother sequence data SM

$$P(SM \mid m, i) = \sum_{m_d} P_{X \mid m_d}(am_i) P(m_d \mid m)$$

where  $P(m_d|m)$  is defined as in previous section and  $P_{X|m_d}$  (am<sub>t</sub>) probability from a binomial distribution is defined as before in the parent data likelihood section. A similar equation applies to the paternal sequence data.

Free floating DNA sequence data dropout:

$$P(S \mid m, \, c, \, H, \, cf, \, i) = \sum_{m_d, c_d} P(S \mid \mu(m_d, \, c_d, \, cf, \, H), \, i) P(m_d \mid m) P(c_d \mid c)$$

where  $P(S|(m_d, c_d, cf, H), i)$  is as defined in the section on free floating data likelihood.

In an embodiment,  $p(m_d|m)$  is the probability of observed mother genotype m, given true mother genotype m, assuming dropout rate  $d_{ps}$ , and  $p(c_d|c)$  is the probability of observed child genotype ca, given true child genotype c, assuming dropout rate  $d_{cs}$ . If  $nA_T$ =number of A alleles in true genotype c,  $nA_D$ =number of A alleles in observed genotype  $c_d$ , where  $nA_T$ <ahref="\sim nA\_D">n and similarly  $nB_T$ </a>=number of B alleles in true genotype c,  $nB_D$ =number of B alleles in observed genotype  $c_d$  where  $nB_T$ </a>>> $nB_D$ </a>=number of B alleles in the genotype  $c_d$  where  $nB_T$ </a>>> $nB_D$ </a> and d</a>=dropout rate, 40 then

$$p(c_d \mid c) = \binom{nA_T}{nA_D} * d^{nA_T - nA_D} * (1-d)^{nA_D} * \binom{nB_T}{nB_D} * d^{nB_T - nB_D} * (1-d)^{nB_D}$$

In an embodiment, the informatics method may incorporate random and consistent bias.

In an ideal word there is no per SNP consistent sampling bias or random noise (in addition to the binomial distribution variation) in the number of sequence counts. In particular, on SNP i, for mother genotype m, true child genotype c and child fraction cf, and X=the number of A's in the set of (A+B) reads on SNP i, X acts like a X-Binomial(p, A+B),  $_{55}$  where p= $\mu$ (m, c, cf, H)=true probability of A content.

In an embodiment, the informatics method may incorporate random bias. As is often the case, suppose that there is a bias in the measurements, so that the probability of getting an A on this SNP is equal to q, which is a bit different than 60 p as defined above. How much different p is from q depends on the accuracy of the measurement process and number of other factors and can be quantified by standard deviations of q away from p. In an embodiment, it is possible to model q as having a beta distribution, with parameters  $\alpha, \beta$  depending 65 on the mean of that distribution being centered at p, and some specified standard deviation s. In particular, this gives

168

XIq~Bin(q,D<sub>i</sub>),where q~Beta( $\alpha$ , $\beta$ ).If we let E(q)=p,V(q)=s<sup>2</sup>, and parameters  $\alpha$ , $\beta$  can be derived as  $\alpha$ =pN,  $\beta$ =(1-p)N, where

$$N = \frac{p(1-p)}{s^2} - 1.$$

This is the definition of a beta-binomial distribution, where one is sampling from a binomial distribution with variable parameter q, where q follows a beta distribution with mean p. So, in a setup with no bias, on SNP i, the parent sequence data (SM) probability assuming true mother genotype (m), given mother sequence A count on SNP i (am<sub>i</sub>) and mother sequence B count on SNP i (bm<sub>i</sub>) may be calculated as:

$$P(SM|m,i)=P_{X|m}(am_i)$$
 where  $X|m\sim \mathrm{Binom}(p_m(A),am_i+bm_i)$ 

20 Now, including random bias with standard deviation s, this becomes:

$$X|m\sim \text{BetaBinom}(p_m(A),am_i+bm_i,s)$$

In the case with no bias, the maternal plasma DNA sequence data (S) probability assuming true mother genotype (m), true child genotype (c), child fraction (cf), assuming child hypothesis H, given free floating DNA sequence A count on SNP i (a,) and free floating sequence B count on SNP i (b,) may be calculated as

$$P(S|m,c,cf,H,i)=P_x(a_i)$$

where X~Binom(p(A),  $a_i+b_i$ ) with p(A)= $\mu$ (m, c, cf, H).

In an embodiment, including random bias with standard deviation s, this becomes X-BetaBinom(p(A),a<sub>i</sub>+b<sub>i</sub>,s), where the amount of extra variation is specified by the deviation parameter s, or equivalently N. The smaller the value of s (or the larger the value of N) the closer this distribution is to the regular binomial distribution. It is possible to estimate the amount of bias, i.e. estimate N above, from unambiguous contexts AA|AA, BB|BB, AA|BB, BB|AA and use estimated N in the above probability. Depending on the behavior of the data, N may be made to be a constant irrespective of the depth of read a<sub>i</sub>+b<sub>i</sub>, or a function of a<sub>i</sub>+b<sub>i</sub>, making bias smaller for larger depths of read.

In an embodiment, the informatics method may incorporate consistent per-SNP bias. Due to artifacts of the sequencing process, some SNPs may have consistently lower or higher counts irrespective of the true amount of A content. Suppose that SNP i consistently adds a bias of w<sub>i</sub> percent to the number of A counts. In some embodiments, this bias can be estimated from the set of training data derived under same conditions, and added back in to the parent sequence data estimate as:

$$P(SM|m,i) = P_{X|m}(am_i) \text{ where } X|m \sim \text{BetaBinom}(p_m(A) + w_i, am_i + bm_i, s)$$

and with the free floating DNA sequence data probability estimate as:

$$P(S|m,c,cf,H,i)=P_x(a_i)$$
 where  $X\sim \text{BetaBinom}(p(A)+w_i,a_i+b_i,s)$ ,

In some embodiments, the method may be written to specifically take into account additional noise, differential sample quality, differential SNP quality, and random sampling bias. An example of this is given here. This method has been shown to be particularly useful in the context of data generated using the massively multiplexed mini-PCR pro-

169

tocol, and was used in Examples 7 through 13. The method involves several steps that each introduce different kind of noise and/or bias to the final model:

Suppose the first sample that comprises a mixture of maternal and fetal DNA contains an original amount of 5 DNA of size=N<sub>0</sub> molecules, usually in the range 1,000-40, 000, where p=true % refs

In the amplification using the universal ligation adaptors, assume that  $N_1$  molecules are sampled; usually  $N_1{\sim}N_0/2$  molecules and random sampling bias is introduced due to  $_{10}$  sampling. The amplified sample may contain a number of molecules  $N_2$  where  $N_2{>>}N_1$ . Let  $X_1$  represent the amount of reference loci (on per SNP basis) out of  $N_1$  sampled molecules, with a variation in  $p_1{=}X_1/N_1$  that introduces random sampling bias throughout the rest of protocol. This sampling bias is included in the model by using a Beta-Binomial (BB) distribution instead of using a simple Binomial distribution model. Parameter N of the Beta-Binomial distribution may be estimated later on per sample basis from training data after adjusting for leakage and amplification  $_{10}$  bias, on SNPs with  $0{<}p{<}1$ . Leakage is the tendency for a SNP to be read incorrectly.

The amplification step will amplify any allelic bias, thus amplification bias introduced due to possible uneven amplification. Suppose that one allele at a locus is amplified f 25 times another allele at that locus is amplified g times, where f=ge<sup>b</sup>, where b=0 indicates no bias. The bias parameter, b, is centered at 0, and indicates how much more or less the A allele get amplified as opposed to the B allele on a particular SNP. The parameter b may differ from SNP to SNP. Bias 30 parameter b may be estimated on per SNP basis, for example from training data.

The sequencing step involves sequencing a sample of amplified molecules. In this step there may be leakage, where leakage is the situation where a SNP is read incorrectly. Leakage may result from any number of problems, and may result in a SNP being read not as the correct allele A, but as another allele B found at that locus or as an allele C or D not typically found at that locus. Suppose the sequencing measures the sequence data of a number of DNA 40 molecules from an amplified sample of size N3, where N3<N2. In some embodiments, N3 may be in the range of 20,000 to 100,000; 100,000 to 500,000; 500,000 to 4,000, 000; 4,000,000 to 20,000,000; or 20,000,000 to 100,000, 000. Each molecule sampled has a probability  $p_g$  of being 45 read correctly, in which case it will show up correctly as allele A. The sample will be incorrectly read as an allele unrelated to the original molecule with probability 1-pg, and will look like allele A with probability pr, allele B with probability  $p_m$  or allele C or allele D with probability  $p_o$ , 50 where  $p_r + p_m + p_o = 1$ . Parameters  $p_g$ ,  $p_r$ ,  $p_m$ ,  $p_o$  are estimated on per SNP basis from the training data.

Different protocols may involve similar steps with variations in the molecular biology steps resulting in different amounts of random sampling, different levels of amplification and different leakage bias. The following model may be equally well applied to each of these cases. The model for the amount of DNA sampled, on per SNP basis, is given by:

$$X_3$$
~BetaBinomial( $L(F(p,b),p_r,p_g),N*H(p,b)$ )

where p=the true amount of reference DNA, b=per SNP bias, and as described above,  $p_g$  is the probability of a correct read,  $p_r$  is the probability of read being read incorrectly but serendipitously looking like the correct allele, in case of a bad read, as described above, and:

$$\begin{split} F(p,b) = & pe^b/(pe^b + (1-p)), & H(p,b) = (e^bp + (1-p))^2/e^b, \\ & L(p,p_{rp}g) = p *p_g + p_r * (1-p_g). \end{split}$$

170

In some embodiments, the method uses a Beta-Binomial distribution instead of a simple binomial distribution; this takes care of the random sampling bias. Parameter N of the Beta-Binomial distribution is estimated on per sample basis on an as needed basis. Using bias correction F(p,b), H(p,b), instead of just p, takes care of the amplification bias. Parameter b of the bias is estimated on per SNP basis from training data ahead of time.

In some embodiments the method uses leakage correction  $L(p,p_{r},p_{g})$ , instead of just p; this takes care of the leakage bias, i.e. varying SNP and sample quality. In some embodiments, parameters  $p_{g}$ ,  $p_{r}$ ,  $p_{o}$  are estimated on per SNP basis from the training data ahead of time. In some embodiments, the parameters  $p_{g}$ ,  $p_{r}$ ,  $p_{o}$  may be updated with the current sample on the go, to account for varying sample quality.

The model described herein is quite general and can account for both differential sample quality and differential SNP quality. Different samples and SNPs are treated differently, as exemplified by the fact that some embodiments use Beta-Binomial distributions whose mean and variance are a function of the original amount of DNA, as well as sample and SNP quality.

Platform Modeling

Consider a single SNP where the expected allele ratio present in the plasma is r (based on the maternal and fetal genotypes). The expected allele ratio is defined as the expected fraction of A alleles in the combined maternal and fetal DNA. For maternal genotype  $g_m$  and child genotype  $g_c$ , the expected allele ratio is given by equation 1, assuming that the genotypes are represented as allele ratios as well.

$$r = fg_c + (1 - f)g_m \tag{1}$$

The observation at the SNP consists of the number of mapped reads with each allele present,  $\mathbf{n}_a$  and  $\mathbf{n}_b$ , which sum to the depth of read d. Assume that thresholds have already been applied to the mapping probabilities and phred scores such that the mappings and allele observations can be considered correct. A phred score is a numerical measure that relates to the probability that a particular measurement at a particular base is wrong. In an embodiment, where the base has been measured by sequencing, the phred score may be calculated from the ratio of the dye intensity corresponding to the called base to the dye intensity of the other bases. The simplest model for the observation likelihood is a binomial distribution which assumes that each of the d reads is drawn independently from a large pool that has allele ratio r. Equation 2 describes this model.

$$P(n_a, n_b \mid r) = p_{bino}(n_a; n_a + n_b, r) = \binom{n_a + n_b}{n_a} r^{n_a} (1 - r)^{n_b}$$
 (2)

The binomial model can be extended in a number of ways.

When the maternal and fetal genotypes are either all A or all B, the expected allele ratio in plasma will be 0 or 1, and the binomial probability will not be well-defined. In practice, unexpected alleles are sometimes observed in practice. In an embodiment, it is possible to use a corrected allele ratio  $\hat{r}=1/(n_{\alpha}+n_{b})$  to allow a small number of the unexpected allele. In an embodiment, it is possible to use training data to model the rate of the unexpected allele appearing on each SNP, and use this model to correct the expected allele ratio. When the expected allele ratio is not 0 or 1, the observed allele ratio may not converge with a sufficiently high depth of read to the expected allele ratio due to amplification bias or other phenomena. The allele ratio can then be modeled as

171

a beta distribution centered at the expected allele ratio, leading to a beta-binomial distribution for  $P(n_a, n_b|r)$  which has higher variance than the binomial.

The platform model for the response at a single SNP will be defined as  $F(a, b, g_c, g_m, f)$  (3), or the probability of observing  $n_a$ =a and  $n_b$ =b given the maternal and fetal genotypes, which also depends on the fetal fraction through equation 1. The functional form of F may be a binomial distribution, beta-binomial distribution, or similar functions as discussed above

$$F(a,b,g_{c},g_{m}f)=P(n_a=a,n_b=b|g_{c},g_{m}f)=P(n_a=a,n_b=b|r)$$
 
$$(g_{c},g_{m}f)) \qquad (3)$$

In an embodiment, the child fraction may be determined as follows. A maximum likelihood estimate of the fetal fraction f for a prenatal test may be derived without the use of paternal information. This may be relevant where the paternal genetic data is not available, for example where the father of record is not actually the genetic father of the fetus. The fetal fraction is estimated from the set of SNPs where 20 the maternal genotype is 0 or 1, resulting in a set of only two possible fetal genotypes. Define So as the set of SNPs with maternal genotype 0 and  $S_1$  as the set of SNPs with maternal genotype 1. The possible fetal genotypes on  $S_0$  are 0 and 0.5, resulting in a set of possible allele ratios  $R_0(f) = \{0, f/2\}$ . 25 Similarly,  $R_1(f) = \{1-f/2, 1\}$ . This method can be trivially extended to include SNPs where maternal genotype is 0.5, but these SNPs will be less informative due to the larger set of possible allele ratios.

Define  $N_{a0}$  and  $N_{b0}$  as the vectors formed by  $n_{as}$  and  $n_{bs}$  30 for SNPs s in  $S_0$ , and  $N_{a1}$  and  $N_{b1}$  similarly for  $S_1$ . The maximum likelihood estimate f of f is defined by equation 4.

$$\hat{f} = arg \max_{f} P(N_{a0}, N_{b0}|f) P(N_{a1}, N_{b1}|f)$$
 (4)

dent conditioned on the SNP's plasma allele ratio, the probabilities can be expressed as products over the SNPs in each set (5).

$$P(N_{a0},N_{b0}|f=\Pi_{s\in S_0}P(n_{as},n_{bs}|f)$$

$$P(N_{a0}, N_{b1}|f = \prod_{s \in S_1} P(n_{as}, n_{bs}|f)$$
(5)

The dependence on f is through the sets of possible allele ratios  $R_0(f)$  and  $R_1(f)$ . The SNP probability  $P(n_{as}, n_{bs}|f)$  can be approximated by assuming the maximum likelihood 45 genotype conditioned on f. At reasonably high fetal fraction and depth of read, the selection of the maximum likelihood genotype will be high confidence. For example, at fetal fraction of 10 percent and depth of read of 1000, consider a SNP where the mother has genotype zero. The expected  $\,^{50}$ allele ratios are 0 and 5 percent, which will be easily distinguishable at sufficiently high depth of read. Substitution of the estimated child genotype into equation 5 results in the complete equation (6) for the fetal fraction estimate.

$$\begin{split} \hat{f} &= \\ &\arg\max_{f} \left[ \Pi_{s \in S_0} \max_{\left(r_s \in R_0(f)\right)} P(n_{as}, n_{bs} \mid r_s) \; \Pi_{s \in S_1} \max_{\left(r_s \in R_1(f)\right)} P(n_{as}, n_{bs} \mid r_s) \right] \\ &= \\ &60 \end{split}$$

The fetal fraction must be in the range [0, 1] and so the optimization can be easily implemented by a constrained one-dimensional search.

In the presence of low depth of read or high noise level, 65 it may be preferable not to assume the maximum likelihood genotype, which may result in artificially high confidences.

172

Another method would be to sum over the possible genotypes at each SNP, resulting in the following expression (7) for  $P(n_a, n_b|f)$  for a SNP in  $S_0$ . The prior probability P(r)could be assumed uniform over R<sub>0</sub>(f), or could be based on population frequencies. The extension to group  $S_1$  is trivial.

$$P(n_a, n_b|f) = \sum_{r \in R_0(f)} P(n_a, n_a|r) P(r)$$

$$(7)$$

In some embodiments the probabilities may be derived as follows. A confidence can be calculated from the data likelihoods of the two hypotheses Ht and Hf. The likelihood of each hypothesis is derived based on the response model, the estimated fetal fraction, the mother genotypes, allele population frequencies, and the plasma allele counts.

Define the following notation:

 $G_m$ ,  $G_c$  true maternal and child genotypes

Gap Gt true genotypes of alleged father and of true father  $G(g_c, g_m, g_{tf}) = P(G_c = g_c | G_m = g_m, G_{tf} = g_{tf})$  inheritance probabilities

 $P(g)=P(G_{tf}=g)$  population frequency of genotype g at particular SNP

Assuming that the observation at each SNP is independent conditioned on the plasma allele ratio, the likelihood of a paternity hypothesis is the product of the likelihoods on the SNPs. The following equations derive the likelihood for a single SNP. Equation 8 is a general expression for the likelihood of any hypothesis h, which will then be broken down into the specific cases of H, and H<sub>c</sub>

$$\begin{split} &P(n_{cn}n_b|h,G_m,G_{gh}f) = & \Sigma_{g,c} \in (0,0,5,1) P(n_{cn}n_b|G_c = g_{cr}G_m,G_{gh}f) \\ & h_d)P(G_c = g_{cr}G_m,G_{gh}h_d) = & \Sigma_{g,c} \in (0,0,5,1) P(n_{cr}f) \\ & n_b|G_c = & g_{cr}G_md)P(G_c = g_{cr}G_mG_{gh}h) \end{split}$$

$$= \sum_{g_c \in (0,0,5,1)} F(n_a, n_b, g_c, g_m, f) P(G_c = g_c | G_m, G_{tf}, h)$$
(8)

In the case of Ht, the alleged father is the true father and the fetal genotypes are inherited from the maternal geno-Assuming that the allele counts at each SNP are indepen- 35 types and alleged father genotypes according to equation 9.

$$P(n_{\alpha n}n_{b}|H_{p}G_{mr}G_{pf})=\Sigma_{g_{c}\in(0,0,5,1)}F(n_{\alpha n}n_{b}|g_{c},g_{mf})P$$
  
 $(G_{c}=g_{c}|G_{mr}G_{qf}H_{t})=\Sigma_{g_{c}\in(0,0,5,1)}F(n_{\alpha r}n_{br}g_{c},g_{mf})P$   
 $(g_{c'}G_{m'}G_{qf})$  (9)

In the case of  $H_{p}$  the alleged father is not the true father. The best estimate of the true father genotypes are given by the population frequencies at each SNP. Thus, the probabilities of child genotypes are determined by the known mother genotypes and the population frequencies, as in equation 10.

$$\begin{split} P(n_{a},n_{b} \mid H_{t},G_{m},G_{tf},f) &= \sum_{g_{c} \in (0,0.5,1)} F(n_{a},n_{b},g_{c},g_{m},f) \\ & P(G_{c} = g_{c} \mid G_{m},G_{tf},H_{f}) \\ &= \sum_{g_{c} \in (0,0.5,1)} F(n_{a},n_{b},g_{c},g_{m},f) \\ & P(G_{c} = g_{c} \mid G_{m}) \\ &= \sum_{g_{c} \in (0,0.5,1)} \sum_{g_{tf} \in (0,0.5,1)} F(n_{a},\\ & n_{b},g_{c},g_{m},f) P(G_{c} = g_{c} \mid G_{m},\\ & G_{tf} = g_{tf}) P(G_{tf} = g_{tf}) \\ &= \sum_{g_{c} \in (0,0.5,1)} \sum_{g_{tf} \in (0,0.5,1)} F(n_{a},\\ & n_{b},g_{c},g_{m},f) G(g_{c},G_{m},g_{tf}) P(g_{tf},g_{tf}) \\ &= \sum_{g_{c} \in (0,0.5,1)} F(n_{a},g_{tf},g_{tf},g_{tf}) P(g_{tf},g_{t$$

The confidence C<sub>n</sub> on correct paternity is calculated from the product over SNPs of the two likelihoods using Bayes rule (11).

US 11,519,035 B2

173

$$Cp = \frac{\prod_{s} P(n_{as}, n_{bs} \mid H_{f}, G_{ms}, G_{tf}, f)}{\prod_{s} P(n_{as}, n_{bs} \mid H_{f}, G_{ms}, G_{tf}, f) +}$$

$$\prod_{s} P(n_{as}, n_{bs} \mid H_{f}, G_{ms}, G_{tf}, f)$$
(11)

Exemplary Methods for Identifying and Analyzing Multiple Pregnancies

In some embodiments, any of the methods of the present invention are used to detect the presence of a multiple 10 pregnancy, such as a twin pregnancy, where at least one of the fetuses is genetically different from at least one other fetus. In some embodiments, fraternal twins are identified based on the presence of two fetus with different allele, different allele ratios, or different allele distributions at some 15 (or all) of the tested loci. In some embodiments, fraternal twins are identified by determining the expected allele ratio at each locus (such as SNP loci) for two fetuses that may have the same or different fetal fractions in the sample (such as a plasma sample). In some embodiments, the likelihood 20 of a particular pair of fetal fractions (where fl is the fetal fraction for fetus 1, and f2 is the fetal fraction for fetus 2) is calculated by considering some or all of the possible genotypes of the two fetuses, conditioned on the mother's genotype and genotype population frequencies. The mixture of 25 two fetal and one maternal genotype, combined with the fetal fractions, determine the expected allele ratio at a SNP. For example, if the mother is AA, fetus 1 is AA, and fetus 2 is AB, the overall fraction of B allele at the SNP is one-half of f2. The likelihood calculation asks how well all of the 30 SNPs together match the expected allele ratios based on all of the possible combinations of fetal genotypes. The fetal fraction pair (f1, f2) that best matches the data is selected. It is not necessary to calculated specific genotypes of the fetuses; instead, one can, for example, considered all of the 35 possible genotypes in a statistical combination. In some embodiments, if the method does not distinguish between singleton and identical twins, an ultrasound can be performed to determine whether there is a singleton or identical twin pregnancy. If the ultrasound detects a twin pregnancy 40 it can be assumed that the pregnancy is an identical twin pregnancy because a fraternal twin pregnancy would have been detected based on the SNP analysis discussed above.

In some embodiments, a pregnant mother is known to have a multiple pregnancy (such as a twin pregnancy) based 45 on prior testing, such as an ultrasound. Any of the methods of the present invention can be used to determine whether the multiple pregnancy includes identical or fraternal twins. For example, the measured allele ratios can be compared to what would be expected for identical twins (the same allele 50 ratios as a singleton pregnancy) or for fraternal twins (such as the calculation of allele ratios as described above). Some identical twins are monochorionic twins, which have a risk of twin-to-twin transfusion syndrome. Thus, twins determined to be identical twins using a method of the invention 55 are desirably tested (such as by ultrasound) to determine if they are monochorionic twins, and if so, these twins can be monitored (such as bi-weekly ultrasounds from 16 weeks) for signs of win-to-twin transfusion syndrome.

In some embodiments, any of the methods of the present 60 invention are used to determine whether any of the fetuses in a multiple pregnancy, such as a twin pregnancy, are aneuploid. Aneuploidy testing for twins begins with the fetal fraction estimate. In some embodiments, the fetal fraction pair (f1, f2) that best matches the data is selected as 65 described above. In some embodiments, a maximum likelihood estimate is performed for the parameter pair (f1, f2)

174

over the range of possible fetal fractions. In some embodiments, the range of f2 is from 0 to f1 because f2 is defined as the smaller fetal fraction. Given a pair (f1, f2), data likelihood is calculated from the allele ratios observed at a set of loci such as SNP loci. In some embodiments, the data likelihood reflects the genotypes of the mother, the father if available, population frequencies, and the resulting probabilities of fetal genotypes. In some embodiments, SNPs are assumed independent. The estimated fetal fraction pair is the one that produces the highest data likelihood. If f2 is 0 then the data is best explained by only one set of fetal genotypes, indicating identical twins, where f1 is the combined fetal fraction. Otherwise f1 and f2 are the estimates of the individual twin fetal fractions. Having established the best estimate of (f1, f2), one can predict the overall fraction of B allele in the plasma for any combination of maternal and fetal genotypes, if desired. It is not necessary to assign individual sequence reads to the individual fetuses. Ploidy testing is performed using another maximum likelihood estimate which compares the data likelihood of two hypotheses. In some embodiments for identical twins, one consider the hypotheses (i) both twins are euploid, and (ii) both twins are trisomic. In some embodiments for fraternal twins, one considers the hypotheses (i) both twins are euploid and (ii) at least one twin is trisomic. The trisomy hypotheses for fraternal twins are based on the lower fetal fraction, since a trisomy in the twin with a higher fetal fraction would also be detected. Ploidy likelihoods are calculated using a method which predicts the expected number of reads at each targeted genome locus conditioned on either the disomy or trisomy hypothesis. There is no requirement for a disomy reference chromosome. The variance model for the expected number of reads takes into account the performance of individual target loci as well as the correlation between loci (see, for example, U.S. Ser. No. 62/008,235, filed Jun. 5, 2014, and U.S. Ser. No. 62/032,785, filed Aug. 4, 2014, which are each hereby incorporated by reference in its entirety). If the smaller twin has fetal fraction f1, our ability to detect a trisomy in that twin is equivalent to our ability to detect a trisomy in a singleton pregnancy at the same fetal fraction. This is because the part of the method that detects the trisomy in some embodiments does not depend on genotypes and does not distinguish between multiple or singleton pregnancy. It simply looks for an increased number of reads in accordance with the determined fetal fraction.

In some embodiments, the method includes detecting the presence of twins based on SNP loci (such as described above). If twins are detected, SPNs are used to determine the fetal fraction of each fetus (f1, f2) such as described above. In some embodiments, samples that have high confidence disomy calls are used to determine the amplification bias on a per-SNP basis. In some embodiments, these samples with high confidence disomy calls are analyzed in the same run as one or more samples of interest. In some embodiments, the amplification bias on a per-SNP basis is used to model the distribution of reads for one or more chromosomes or chromosome segments of interest such as chromosome 21 that are expected or the disomy hypothesis and the trisomy hypothesis given the lower of the two twin fetal fraction. The likelihood or probability of disomy or trisomy is calculated given the two models and the measured quantity of the chromosome or chromosome segment of interest.

In some embodiments, the threshold for a positive aneuploidy call (such as a trisomy call) is set based on the twin with the lower fetal fraction. This way, if the other twin is positive, or if both are positive, the total chromosome representation is definitely above the threshold.

## US 11,519,035 B2

175

Maximum Likelihood Model using Percent Fetal Fraction Determining the ploidy status of a fetus by measuring the free floating DNA contained in maternal serum, or by measuring the genotypic material in any mixed sample, is a non-trivial exercise. There are a number of methods, for example, performing a read count analysis where the presumption is that if the fetus is trisomic at a particular chromosome, then the overall amount of DNA from that chromosome found in the maternal blood will be elevated with respect to a reference chromosome. One way to detect 10 trisomy in such fetuses is to normalize the amount of DNA expected for each chromosome, for example, according to the number of SNPs in the analysis set that correspond to a given chromosome, or according to the number of uniquely mappable portions of the chromosome. Once the measure- 15 ments have been normalized, any chromosomes for which the amount of DNA measured exceeds a certain threshold are determined to be trisomic. This approach is described in Fan, et al. PNAS, 2008; 105(42); pp. 16266-16271, and also in Chiu et al. BMJ 2011; 342:c7401. In the Chiu et al. paper, 20 the normalization was accomplished by calculating a Z score as follows:

Z score for percentage chromosome 21 in test case– ((percentage chromosome 21 in test case)– (mean percentage chromosome 21 in reference controls))/(standard deviation of percentage chromosome 21 in reference controls).

These methods determine the ploidy status of the fetus using a single hypothesis rejection method. However, they suffer from some significant shortcomings. Since these 30 methods for determining ploidy in the fetus are invariant according to the percentage of fetal DNA in the sample, they use one cut off value; the result of this is that the accuracies of the determinations are not optimal, and those cases where the percentage of fetal DNA in the mixture are relatively low 35 will suffer the worst accuracies.

In an embodiment, a method of the present disclosure is used to determine the ploidy state of the fetus involves taking into account the fraction of fetal DNA in the sample. In another embodiment of the present disclosure, the method 40 involves the use of maximum likelihood estimations. In an embodiment, a method of the present disclosure involves calculating the percent of DNA in a sample that is fetal or placental in origin. In an embodiment, the threshold for calling aneuploidy is adaptively adjusted based on the 45 calculated percent fetal DNA. In some embodiments, the method for estimating the percentage of DNA that is of fetal origin in a mixture of DNA, comprises obtaining a mixed sample that comprises genetic material from the mother, and genetic material from the fetus, obtaining a genetic sample 50 from the father of the fetus, measuring the DNA in the mixed sample, measuring the DNA in the father sample, and calculating the percentage of DNA that is of fetal origin in the mixed sample using the DNA measurements of the mixed sample, and of the father sample.

In an embodiment of the present disclosure, the fraction of fetal DNA, or the percentage of fetal DNA in the mixture can be measured. In some embodiments the fraction can be calculated using only the genotyping measurements made on the maternal plasma sample itself, which is a mixture of fetal and maternal DNA. In some embodiments the fraction may be calculated also using the measured or otherwise known genotype of the mother and/or the measured or otherwise known genotype of the father. In some embodiments the percent fetal DNA may be calculated using the measurements made on the mixture of maternal and fetal DNA along with the knowledge of the parental contexts. In an embodi-

176

ment, the fraction of fetal DNA may be calculated using population frequencies to adjust the model on the probability on particular allele measurements.

In an embodiment of the present disclosure, a confidence may be calculated on the accuracy of the determination of the ploidy state of the fetus. In an embodiment, the confidence of the hypothesis of greatest likelihood  $(H_{major})$  may be calculated as  $(1-H_{major})/\Sigma(\text{all H})$ . It is possible to determine the confidence of a hypothesis if the distributions of all of the hypotheses are known. It is possible to determine the distribution of all of the hypotheses if the parental genotype information is known. It is possible to calculate a confidence of the ploidy determination if the knowledge of the expected distribution of data for the euploid fetus and the expected distribution of data for the aneuploid fetus are known. It is possible to calculate these expected distributions if the parental genotype data are known. In an embodiment one may use the knowledge of the distribution of a test statistic around a normal hypothesis and around an abnormal hypothesis to determine both the reliability of the call as well as refine the threshold to make a more reliable call. This is particularly useful when the amount and/or percent of fetal DNA in the mixture is low. It will help to avoid the situation where a fetus that is actually aneuploid is found to be 25 euploid because a test statistic, such as the Z statistic does not exceed a threshold that is made based on a threshold that is optimized for the case where there is a higher percent fetal DNA.

In an embodiment, a method disclosed herein can be used to determine a fetal aneuploidy by determining the number of copies of maternal and fetal target chromosomes in a mixture of maternal and fetal genetic material. This method may entail obtaining maternal tissue comprising both maternal and fetal genetic material; in some embodiments this maternal tissue may be maternal plasma or a tissue isolated from maternal blood. This method may also entail obtaining a mixture of maternal and fetal genetic material from said maternal tissue by processing the aforementioned maternal tissue. This method may entail distributing the genetic material obtained into a plurality of reaction samples, to randomly provide individual reaction samples that comprise a target sequence from a target chromosome and individual reaction samples that do not comprise a target sequence from a target chromosome, for example, performing high throughput sequencing on the sample. This method may entail analyzing the target sequences of genetic material present or absent in said individual reaction samples to provide a first number of binary results representing presence or absence of a presumably euploid fetal chromosome in the reaction samples and a second number of binary results representing presence or absence of a possibly aneuploid fetal chromosome in the reaction samples. Either of the number of binary results may be calculated, for example, by way of an informatics technique that counts sequence reads that map to a particular chromosome, to a particular region of a chromosome, to a particular locus or set of loci. This method may involve normalizing the number of binary events based on the chromosome length, the length of the region of the chromosome, or the number of loci in the set. This method may entail calculating an expected distribution of the number of binary results for a presumably euploid fetal chromosome in the reaction samples using the first number. This method may entail calculating an expected distribution of the number of binary results for a presumably aneuploid fetal chromosome in the reaction samples using the first number and an estimated fraction of fetal DNA found in the mixture, for example, by multiplying the

## US 11,519,035 B2

177

expected read count distribution of the number of binary results for a presumably euploid fetal chromosome by (1+n/2) where n is the estimated fetal fraction. In some embodiments, the sequence reads may be treated at probabilistic mappings rather than binary results; this method 5 would yield higher accuracies, but require more computing power. The fetal fraction may be estimated by a plurality of methods, some of which are described elsewhere in this disclosure. This method may involve using a maximum likelihood approach to determine whether the second number corresponds to the possibly aneuploid fetal chromosome being euploid or being aneuploid. This method may involve calling the ploidy status of the fetus to be the ploidy state that corresponds to the hypothesis with the maximum likelihood of being correct given the measured data.

Note that the use of a maximum likelihood model may be used to increase the accuracy of any method that determines the ploidy state of a fetus. Similarly, a confidence maybe calculated for any method that determines the ploidy state of the fetus. The use of a maximum likelihood model would 20 result in an improvement of the accuracy of any method where the ploidy determination is made using a single hypothesis rejection technique. A maximum likelihood model may be used for any method where a likelihood distribution can be calculated for both the normal and 25 abnormal cases. The use of a maximum likelihood model implies the ability to calculate a confidence for a ploidy call.

Further Discussion of the Method

In an embodiment, a method disclosed herein utilizes a quantitative measure of the number of independent obser- 30 vations of each allele at a polymorphic locus, where this does not involve calculating the ratio of the alleles. This is different from methods, such as some microarray based methods, which provide information about the ratio of two alleles at a locus but do not quantify the number of inde- 3: pendent observations of either allele. Some methods known in the art can provide quantitative information regarding the number of independent observations, but the calculations leading to the ploidy determination utilize only the allele ratios, and do not utilize the quantitative information. To 40 illustrate the importance of retaining information about the number of independent observations consider the sample locus with two alleles, A and B. In a first experiment twenty A alleles and twenty B alleles are observed, in a second experiment 200 A alleles and 200 B alleles are observed. In 45 both experiments the ratio (A/(A+B)) is equal to 0.5, however the second experiment conveys more information than the first about the certainty of the frequency of the A or B allele. The instant method, rather than utilizing the allele ratios, uses the quantitative data to more accurately model 50 the most likely allele frequencies at each polymorphic locus.

In an embodiment, the instant methods build a genetic model for aggregating the measurements from multiple polymorphic loci to better distinguish trisomy from disomy and also to determine the type of trisomy. Additionally, the 5. instant method incorporates genetic linkage information to enhance the accuracy of the method. This is in contrast to some methods known in the art where allele ratios are averaged across all polymorphic loci on a chromosome. The method disclosed herein explicitly models the allele fre- 60 quency distributions expected in disomy as well as and trisomy resulting from nondisjunction during meiosis I, nondisjunction during meiosis II, and nondisjunction during mitosis early in fetal development. To illustrate why this is important, if there were no crossovers nondisjunction during 65 meiosis I would result a trisomy in which two different homologs were inherited from one parent; nondisjunction

178

during meiosis II or during mitosis early in fetal development would result in two copies of the same homolog from one parent. Each scenario results in different expected allele frequencies at each polymorphic locus and also at all physically linked loci (i.e. loci on the same chromosome) considered jointly. Crossovers, which result in the exchange of genetic material between homologs, make the inheritance pattern more complex, but the instant method accommodates for this by using genetic linkage information, i.e. recombination rate information and the physical distance between loci. To better distinguish between meiosis I nondisjunction and meiosis II or mitotic nondisjunction the instant method incorporates into the model an increasing probability of crossover as the distance from the centromere increases. Meiosis II and mitotic nondisjunction can distinguished by the fact that mitotic nondisjunction typically results in identical or nearly identical copies of one homolog while the two homologs present following a meiosis II nondisjunction event often differ due to one or more crossovers during gametogenesis.

In an embodiment, a method of the present disclosure may not determine the haplotypes of the parents if disomy is assumed. In an embodiment, in case of trisomy, the instant method can make a determination about the haplotypes of one or both parents by using the fact that plasma takes two copies from one parent, and parent phase information can be determined by noting which two copies have been inherited from the parent in question. In particular, a child can inherit either two of the same copies of the parent (matched trisomy) or both copies of the parent (unmatched trisomy). At each SNP one can calculate the likelihood of the matched trisomy and of the unmatched trisomy. A ploidy calling method that does not use the linkage model accounting for crossovers would calculate the overall likelihood of the trisomy as a simple weighted average of the matched and unmatched trisomies over all chromosomes. However, due to the biological mechanisms that result in disjunction error and crossing over, trisomy can change from matched to unmatched (and vice versa) on a chromosome only if a crossover occurs. The instant method probabilistically takes into account the likelihood of crossover, resulting in ploidy calls that are of greater accuracy than those methods that do not.

In an embodiment, a reference chromosome is used to determine the child fraction and noise level amount or probability distribution. In an embodiment, the child fraction, noise level, and/or probability distribution is determined using only the genetic information available from the chromosome whose ploidy state is being determined. The instant method works without the reference chromosome, as well as without fixing the particular child fraction or noise level. This is a significant improvement and point of differentiation from methods known in the art where genetic data from a reference chromosome is necessary to calibrate the child fraction and chromosome behavior.

In an embodiment where a reference chromosome is not needed to determine the fetal fraction, determining the hypothesis is done as follows:

 $H^* = \operatorname*{argmax}_u LIK(D \mid H) * priorprob(H)$ 

With the algorithm with reference chromosome, one typically assumes that the reference chromosome is a disomy, and then one may either (a) fix the most likely child

179

fraction and random noise level N based on this assumption and reference chromosome data:

$$[\mathit{cfr}^*,\,N^*] = \underset{\mathit{cfr},N}{\operatorname{argmax}}\,\mathit{LIK}(D(\mathit{ref.chrom}) \mid H11,\,\mathit{cfr},\,N)$$

And then reduce

$$LIK(D|H)=LIK(D|H,cfr^*,N^*)$$

or (b) estimate the child fraction and noise level distribution based on this assumption and reference chromosome data. In particular, one would not fix just one value for cfr and N, but assign probability p(cfr, N) for the wider range of possible cfr. N values:

$$p(cfr,N)\sim LIK(D(ref.chrom)H11,cfr,N)*priorprob(cfr,N)$$

where priorprob(cfr, N) is the prior probability of particular child fraction and noise level, determined by prior knowledge and experiments. If desired, just uniform over the range of cfr, N. One may then write:

$$LIK(D \mid H) = \sum_{cfr,N} LIK(D \mid H, \, cfr, \, N) * p(cfr, \, N)$$

Both methods above give good results.

Note that in some instances using a reference chromosome is not desirable, possible or feasible. In such a case, it is possible to derive the best ploidy call for each chromosome separately. In particular:

$$LIK(D \mid H) = \sum_{cfr,N} LIK(D \mid H, cfr, N) * p(cfr, N \mid H)$$

p(cfr, N|H) may be determined as above, for each chromosome separately, assuming hypothesis H, not just for the 40 reference chromosome assuming disomy. It is possible, using this method, to keep both noise and child fraction parameters fixed, fix either of the parameters, or keep both parameters in probabilistic form for each chromosome and each hypothesis.

Measurements of DNA are noisy and/or error prone, especially measurements where the amount of DNA is small, or where the DNA is mixed with contaminating DNA. This noise results in less accurate genotypic data, and less accurate ploidy calls. In some embodiments, platform modeling or some other method of noise modeling may be used to counter the deleterious effects of noise on the ploidy determination. The instant method uses a joint model of both channels, which accounts for the random noise due to the amount of input DNA, DNA quality, and/or protocol quality.

This is in contrast to some methods known in the art where the ploidy determinations are made using the ratio of allele intensities at a locus. This method precludes accurate SNP noise modeling. In particular, errors in the measurements typically do not specifically depend on the measured channel intensity ratio, which reduces the model to using one-dimensional information. Accurate modeling of noise, channel quality and channel interaction requires a two-dimensional joint model, which can not be modeled using allele ratios.

In particular, projecting two channel information to the ratio r where f(x,y) is r=x/y, does not lend itself to accurate

180

channel noise and bias modeling. Noise on a particular SNP is not a function of the ratio, i.e. noise(x,y) f(x,y) but is in fact a joint function of both channels. For example, in the binomial model, noise of the measured ratio has a variance of r(1-r)/(x+y) which is not a function purely of r. In such a model, where any channel bias or noise is included, suppose that on SNP i, the observed channel X value is  $x=a_iX+b_i$ , where X is the true channel value,  $b_i$  is the extra channel bias and random noise. Similarly, suppose that  $y=c_iY+d_i$ . The observed ratio r-x/y can not accurately predict the true ratio X/Y or model the leftover noise, since (aiX+bi)/(ciY+di) is not a function of X/Y.

The method disclosed herein describes an effective way to model noise and bias using joint binomial distributions of all of the measurement channels individually. Relevant equations may be found elsewhere in the document in sections which speaks of per SNP consistent bias, P(good) and P(reflbad), P(mutlbad) which effectively adjust SNP behavior. In an embodiment, a method of the present disclosure uses a BetaBinomial distribution, which avoids the limiting practice of relying on the allele ratios only, but instead models the behavior based on both channel counts.

In an embodiment, a method disclosed herein can call the ploidy of a gestating fetus from genetic data found in maternal plasma by using all available measurements. In an embodiment, a method disclosed herein can call the ploidy of a gestating fetus from genetic data found in maternal plasma by using the measurements from only a subset of parental contexts. Some methods known in the art only use measured genetic data where the parental context is from the AA|BB context, that is, where the parents are both homozygous at a given locus, but for a different allele. One problem with this method is that a small proportion of polymorphic loci are from the AA|BB context, typically less than 10%. In an embodiment of a method disclosed herein, the method does not use genetic measurements of the maternal plasma made at loci where the parental context is AA|BB. In an embodiment, the instant method uses plasma measurements for only those polymorphic loci with the AA|AB, AB|AA, and AB|AB parental context.

Some methods known in the art involve averaging allele ratios from SNPs in the AAIBB context, where both parent genotypes are present, and claim to determine the ploidy calls from the average allele ratio on these SNPs. This method suffers from significant inaccuracy due differential SNP behavior. Note that this method assumes that have both parent genotypes are known. In contrast, in some embodiments, the instant method uses a joint channel distribution model that does not assume the presence of either of the parents, and does not assume the uniform SNP behavior. In some embodiments, the instant method accounts for the different SNP behavior/weighing. In some embodiments, the instant method does not require the knowledge of one or both parental genotypes. An example of how the instant method may accomplish this follows:

In some embodiments, the log likelihood of a hypothesis may be determined on a per SNP basis. On a particular SNP i, assuming fetal ploidy hypothesis H and percent fetal DNA cf, the log likelihood of observed data D is defined as:

$$LIK(D \mid H, i) = \log P(D \mid H, cf, i) = \log \left( \sum_{m, f, c} P(D \mid m, f, c, H, cf, i) P(c \mid m, f, H) P(m \mid i) P(f \mid i) \right)$$

## US 11,519,035 B2

18

where m are possible true mother genotypes, f are possible true father genotypes, where  $m,f \in \{AA,AB,BB\}$ , and where c are possible child genotypes given the hypothesis H. In particular, for monosomy c  $\{A,B\}$ , for disomy c $\in \{AA,AB,BB\}$ , for trisomy c  $\in \{AAA,AAB,ABB,BBB\}$ . Note that including parental genotypic data typically results in more accurate ploidy determinations, however, parental genotypic data is not necessary for the instant method to work well.

Some methods known in the art involve averaging allele 10 ratios from SNPs where the mother is homozygous but a different allele is measured in the plasma (either AA|AB or AA|BB contexts), and claim to determine the ploidy calls from the average allele ratio on these SNPs. This method is intended for cases where the paternal genotype is not 15 available. Note that it is questionable how accurately one can claim that plasma is heterozygous on a particular SNP without the presence of homozygous and opposite father BB: for cases with low child fraction, what looks like presence of B allele could be just presence of noise; addi- 20 tionally, what looks like no B present could be simple allele drop out of the fetal measurements. Even in a case where one can actually determine heterozygosity of the plasma, this method will not be able to distinguish paternal trisomies. In particular, for SNPs where mother is AA, and where some B 25 is measured in the plasma, if the father is GG, the resulting child genotype is ΔGG, resulting in an average ratio of 33% A (for child fraction=100%). But in the case where the father is  $\Delta G$ , the resulting child genotype could be  $\Delta GG$  for matched trisomy, contributing to the 33% A ratio, or AAG 30 for unmatched trisomy, drawing the average ratio more toward 66% A. Given that many trisomies are on chromosomes with crossovers, the overall chromosome can have anywhere between no unmatched trisomy and all unmatched trisomy, this ratio can vary anywhere between 33-66%. For 35 a plain disomy, the ratio should be around 50%. Without the use of a linkage model or an accurate error model of the average, this method would miss many cases of paternal trisomy. In contrast, the method disclosed herein assigns parental genotype probabilities for each parental genotypic 40 candidate, based on available genotypic information and population frequency, and does not explicitly require parental genotypes. Additionally, the method disclosed herein is able to detect trisomy even in the absence or presence of parent genotypic data, and can compensate by identifying 45 the points of possible crossovers from matched to unmatched trisomy using a linkage model.

Some methods known in the art claim a method for averaging allele ratios from SNPs where neither the maternal or paternal genotype is known, and for determining the 50 ploidy calls from average ratio on these SNPs. However, a method to accomplish these ends is not disclosed. The method disclosed herein is able to make accurate ploidy calls in such a situation, and the reduction to practice is disclosed elsewhere in this document, using a joint probsability maximum likelihood method and optionally utilizes SNP noise and bias models, as well as a linkage model.

Some methods known in the art involve averaging allele ratios and claim to determine the ploidy calls from the average allele ratio at one or a few SNPs. However, such 60 methods do not utilize the concept of linkage. The methods disclosed herein do not suffer from these drawbacks. Using Sequence Length as a Prior to Determine the Origin of DNA

It has been reported that the distribution of length of 65 sequences differ for maternal and fetal DNA, with fetal generally being shorter. In an embodiment of the present

182

disclosure, it is possible to use previous knowledge in the form of empirical data, and construct prior distribution for expected length of both mother(P(X|maternal)) and fetal DNA (P(X|fetal)). Given new unidentified DNA sequence of length x, it is possible to assign a probability that a given sequence of DNA is either maternal or fetal DNA, based on prior likelihood of x given either maternal or fetal. In particular if P(x|maternal)>P(x|fetal), then the DNA sequence can be classified as maternal, with P(x|maternal)= P(x|maternal)/[(P(x|maternal)+P(x|fetal)], and if p(x|maternal) $\langle p(x|fetal)$ , then the DNA sequence can be classified as fetal, P(x|fetal)=P(x|fetal)/[(P(x|maternal)+P(x|fetal)]. In an embodiment of the present disclosure, a distributions of maternal and fetal sequence lengths can be determined that is specific for that sample by considering the sequences that can be assigned as maternal or fetal with high probability, and then that sample specific distribution can be used as the expected size distribution for that sample.

Variable Read Depth to Minimize Sequencing Cost

In many clinical trials concerning a diagnostic, for example, in Chiu et al. BMJ 2011; 342:c7401, a protocol with a number of parameters is set, and then the same protocol is executed with the same parameters for each of the patients in the trial. In the case of determining the ploidy status of a fetus gestating in a mother using sequencing as a method to measure genetic material one pertinent parameter is the number of reads. The number of reads may refer to the number of actual reads, the number of intended reads, fractional lanes, full lanes, or full flow cells on a sequencer. In these studies, the number of reads is typically set at a level that will ensure that all or nearly all of the samples achieve the desired level of accuracy. Sequencing is currently an expensive technology, a cost of roughly \$200 per 5 mappable million reads, and while the price is dropping, any method which allows a sequencing based diagnostic to operate at a similar level of accuracy but with fewer reads will necessarily save a considerable amount of money.

The accuracy of a ploidy determination is typically dependent on a number of factors, including the number of reads and the fraction of fetal DNA in the mixture. The accuracy is typically higher when the fraction of fetal DNA in the mixture is higher. At the same time, the accuracy is typically higher if the number of reads is greater. It is possible to have a situation with two cases where the ploidy state is determined with comparable accuracies wherein the first case has a lower fraction of fetal DNA in the mixture than the second, and more reads were sequenced in the first case than the second. It is possible to use the estimated fraction of fetal DNA in the mixture as a guide in determining the number of reads necessary to achieve a given level of accuracy.

In an embodiment of the present disclosure, a set of samples can be run where different samples in the set are sequenced to different reads depths, wherein the number of reads run on each of the samples is chosen to achieve a given level of accuracy given the calculated fraction of fetal DNA in each mixture. In an embodiment of the present disclosure, this may entail making a measurement of the mixed sample to determine the fraction of fetal DNA in the mixture; this estimation of the fetal fraction may be done with sequencing, it may be done with TAQMAN, it may be done with qPCR, it may be done with SNP arrays, it may be done with any method that can distinguish different alleles at a given loci. The need for a fetal fraction estimate may be eliminated by including hypotheses that cover all or a selected set of fetal fractions in the set of hypotheses that are considered when comparing to the actual measured data. After the

## US 11,519,035 B2

183

fraction fetal DNA in the mixture has been determined, the number of sequences to be read for each sample may be determined

In an embodiment of the present disclosure, 100 pregnant women visit their respective OB's, and their blood is drawn into blood tubes with an anti-lysant and/or something to inactivate DNAase. They each take home a kit for the father of their gestating fetus who gives a saliva sample. Both sets of genetic materials for all 100 couples are sent back to the laboratory, where the mother blood is spun down and the buffy coat is isolated, as well as the plasma. The plasma comprises a mixture of maternal DNA as well as placentally derived DNA. The maternal buffy coat and the paternal blood is genotyped using a SNP array, and the DNA in the maternal plasma samples are targeted with SURESELECT hybridization probes. The DNA that was pulled down with the probes is used to generate 100 tagged libraries, one for each of the maternal samples, where each sample is tagged with a different tag. A fraction from each library is with- 20 drawn, each of those fractions are mixed together and added to two lanes of a ILLUMINA HISEQ DNA sequencer in a multiplexed fashion, wherein each lane resulted in approximately 50 million mappable reads, resulting in approximately 100 million mappable reads on the 100 multiplexed 25 mixtures, or approximately 1 million reads per sample. The sequence reads were used to determine the fraction of fetal DNA in each mixture. 50 of the samples had more than 15% fetal DNA in the mixture, and the 1 million reads were sufficient to determine the ploidy status of the fetuses with 30 a 99.9% confidence.

Of the remaining mixtures, 25 had between 10 and 15% fetal DNA; a fraction of each of the relevant libraries prepped from these mixtures were multiplexed and run down one lane of the HISEQ generating an additional 2 3 million reads for each sample. The two sets of sequence data for each of the mixture with between 10 and 15% fetal DNA were added together, and the resulting 3 million reads per sample which were sufficient to determine the ploidy state of those fetuses with 99.9% confidence.

Of the remaining mixtures, 13 had between 6 and 10% fetal DNA; a fraction of each of the relevant libraries prepped from these mixtures were multiplexed and run down one lane of the HISEQ generating an additional 4 million reads for each sample. The two sets of sequence data 45 for each of the mixture with between 6 and 10% fetal DNA were added together, and the resulting 5 million total reads per mixture which were sufficient to determine the ploidy state of those fetuses with 99.9% confidence.

Of the remaining mixtures, 8 had between 4 and 6% fetal 50 DNA; a fraction of each of the relevant libraries prepped from these mixtures were multiplexed and run down one lane of the HISEQ generating an additional 6 million reads for each sample. The two sets of sequence data for each of the mixture with between 4 and 6% fetal DNA were added 55 together, and the resulting 7 million total reads per mixture which were sufficient to determine the ploidy state of those fetuses with 99.9% confidence.

Of the remaining four mixtures, all of them had between 2 and 4% fetal DNA; a fraction of each of the relevant 60 libraries prepped from these mixtures were multiplexed and run down one lane of the HISEQ generating an additional 12 million reads for each sample. The two sets of sequence data for each of the mixture with between 2 and 4% fetal DNA were added together, and the resulting 13 million total reads 65 per mixture which were sufficient to determine the ploidy state of those fetuses with 99.9% confidence.

124

This method required six lanes of sequencing on a HISEQ machine to achieve 99.9% accuracy over 100 samples. If the same number of runs had been required for every sample, to ensure that every ploidy determination was made with a 99.9% accuracy, it would have taken 25 lanes of sequencing, and if a no-call rate or error rate of 4% was tolerated, it could have been achieved with 14 lanes of sequencing. Using Raw Genotyping Data

There are a number of methods that can accomplish NPD using fetal genetic information measured on fetal DNA found in maternal blood. Some of these methods involve making measurements of the fetal DNA using SNP arrays, some methods involve untargeted sequencing, and some methods involve targeted sequencing. The targeted sequencing may target SNPs, it may target STRs, it may target other polymorphic loci, it may target non-polymorphic loci, or some combination thereof. Some of these methods may involve using a commercial or proprietary allele caller that calls the identity of the alleles from the intensity data that comes from the sensors in the machine doing the measuring. For example, the ILLUMINA INFINIUM system or the AFFYMETRIX GENECHIP microarray system involves beads or microchips with attached DNA sequences that can hybridize to complementary segments of DNA; upon hybridization, there is a change in the fluorescent properties of the sensor molecule that can be detected. There are also sequencing methods, for example the ILLUMINA SOLEXA GENOME SEQUENCER or the ABI SOLID GENOME SEQUENCER, wherein the genetic sequence of fragments of DNA are sequenced; upon extension of the strand of DNA complementary to the strand being sequenced, the identity of the extended nucleotide is typically detected via a fluorescent or radio tag appended to the complementary nucleotide. In all of these methods the genotypic or sequencing data is typically determined on the basis of fluorescent or other signals, or the lack thereof. These systems are typically combined with low level software packages that make specific allele calls (secondary genetic data) from the analog output of the fluorescent or other detection device (primary genetic data). For example, in the case of a given allele on a SNP array, the software will make a call, for example, that a certain SNP is present or not present if the fluorescent intensity is measure above or below a certain threshold. Similarly, the output of a sequencer is a chromatogram that indicates the level of fluorescence detected for each of the dyes, and the software will make a call that a certain base pair is A or T or C or G. High throughput sequencers typically make a series of such measurements, called a read, that represents the most likely structure of the DNA sequence that was sequenced. The direct analog output of the chromatogram is defined here to be the primary genetic data, and the base pair/SNP calls made by the software are considered here to be the secondary genetic data. In an embodiment, primary data refers to the raw intensity data that is the unprocessed output of a genotyping platform, where the genotyping platform may refer to a SNP array, or to a sequencing platform. The secondary genetic data refers to the processed genetic data, where an allele call has been made, or the sequence data has been assigned base pairs, and/or the sequence reads have been mapped to the genome.

Many higher level applications take advantage of these allele calls, SNP calls and sequence reads, that is, the secondary genetic data, that the genotyping software produces. For example, DNA NEXUS, ELAND or MAQ will take the sequencing reads and map them to the genome. For example, in the context of non-invasive prenatal diagnosis, complex informatics, such as PARENTAL SUPPORT<sup>TM</sup>,

## US 11,519,035 B2

185

may leverage a large number of SNP calls to determine the genotype of an individual. Also, in the context of preimplantation genetic diagnosis, it is possible to take a set of sequence reads that are mapped to the genome, and by taking a normalized count of the reads that are mapped to 5 each chromosome, or section of a chromosome, it may be possible to determine the ploidy state of an individual. In the context of non-invasive prenatal diagnosis it may be possible to take a set of sequence reads that have been measured on DNA present in maternal plasma, and map them to the genome. One may then take a normalized count of the reads that are mapped to each chromosome, or section of a chromosome, and use that data to determine the ploidy state of an individual. For example, it may be possible to conclude that those chromosomes that have a disproportionately large number of reads are trisomic in the fetus that is gestating in the mother from which the blood was drawn.

However, in reality, the initial output of the measuring instruments is an analog signal. When a certain base pair is 20 called by the software that is associated with the sequencing software, for example the software may call the base pair a T, in reality the call is the call that the software believes to be most likely. In some cases, however, the call may be of low confidence, for example, the analog signal may indicate 25 that the particular base pair is only 90% likely to be a T, and 10% likely to be an A. In another example, the genotype calling software that is associated with a SNP array reader may call a certain allele to be G. However, in reality, the underlying analog signal may indicate that it is only 70% likely that the allele is G, and 30% likely that the allele is T. In these cases, when the higher level applications use the genotype calls and sequence calls made by the lower level software, they are losing some information. That is, the 35 primary genetic data, as measured directly by the genotyping platform, may be messier than the secondary genetic data that is determined by the attached software packages, but it contains more information. In mapping the secondary genetic data sequences to the genome, many reads are 40 thrown out because some bases are not read with enough clarity and or mapping is not clear. When the primary genetic data sequence reads are used, all or many of those reads that may have been thrown out when first converted to secondary genetic data sequence read can be used by treat- 45 ing the reads in a probabilistic manner.

In an embodiment of the present disclosure, the higher level software does not rely on the allele calls, SNP calls, or sequence reads that are determined by the lower level software. Instead, the higher level software bases its calcu- 50 lations on the analog signals directly measured from the genotyping platform. In an embodiment of the present disclosure, an informatics based method such as PAREN-TAL SUPPORT<sup>TM</sup> is modified so that its ability to reconstruct the genetic data of the embryo/fetus/child is engi- 55 neered to directly use the primary genetic data as measured by the genotyping platform. In an embodiment of the present disclosure, an informatics based method such as PAREN-TAL SUPPORT<sup>TM</sup> is able to make allele calls, and/or chromosome copy number calls using primary genetic data, 60 and not using the secondary genetic data. In an embodiment of the present disclosure, all genetic calls, SNPs calls, sequence reads, sequence mapping is treated in a probabilistic manner by using the raw intensity data as measured directly by the genotyping platform, rather than converting 65 the primary genetic data to secondary genetic calls. In an embodiment, the DNA measurements from the prepared

186

sample used in calculating allele count probabilities and determining the relative probability of each hypothesis comprise primary genetic data.

In some embodiments, the method can increase the accuracy of genetic data of a target individual which incorporates genetic data of at least one related individual, the method comprising obtaining primary genetic data specific to a target individual's genome and genetic data specific to the genome(s) of the related individual(s), creating a set of one or more hypotheses concerning possibly which segments of which chromosomes from the related individual(s) correspond to those segments in the target individual's genome, determining the probability of each of the hypotheses given the target individual's primary genetic data and the related individual(s)'s genetic data, and using the probabilities associated with each hypothesis to determine the most likely state of the actual genetic material of the target individual. In some embodiments, the method can determining the number of copies of a segment of a chromosome in the genome of a target individual, the method comprising creating a set of copy number hypotheses about how many copies of the chromosome segment are present in the genome of a target individual, incorporating primary genetic data from the target individual and genetic information from one or more related individuals into a data set, estimating the characteristics of the platform response associated with the data set, where the platform response may vary from one experiment to another, computing the conditional probabilities of each copy number hypothesis, given the data set and the platform response characteristics, and determining the copy number of the chromosome segment based on the most probable copy number hypothesis. In an embodiment, a method of the present disclosure can determine a ploidy state of at least one chromosome in a target individual, the method comprising obtaining primary genetic data from the target individual and from one or more related individuals, creating a set of at least one ploidy state hypothesis for each of the chromosomes of the target individual, using one or more expert techniques to determine a statistical probability for each ploidy state hypothesis in the set, for each expert technique used, given the obtained genetic data, combining, for each ploidy state hypothesis, the statistical probabilities as determined by the one or more expert techniques, and determining the ploidy state for each of the chromosomes in the target individual based on the combined statistical probabilities of each of the ploidy state hypotheses. In an embodiment, a method of the present disclosure can determine an allelic state in a set of alleles, in a target individual, and from one or both parents of the target individual, and optionally from one or more related individuals, the method comprising obtaining primary genetic data from the target individual, and from the one or both parents, and from any related individuals, creating a set of at least one allelic hypothesis for the target individual, and for the one or both parents, and optionally for the one or more related individuals, where the hypotheses describe possible allelic states in the set of alleles, determining a statistical probability for each allelic hypothesis in the set of hypotheses given the obtained genetic data, and determining the allelic state for each of the alleles in the set of alleles for the target individual, and for the one or both parents, and optionally for the one or more related individuals, based on the statistical probabilities of each of the allelic hypotheses.

In some embodiments, the genetic data of the mixed sample may comprise sequence data wherein the sequence data may not uniquely map to the human genome. In some embodiments, the genetic data of the mixed sample may

## US 11,519,035 B2

187

comprise sequence data wherein the sequence data maps to a plurality of locations in the genome, wherein each possible mapping is associated with a probability that the given mapping is correct. In some embodiments, the sequence reads are not assumed to be associated with a particular position in the genome. In some embodiments, the sequence reads are associated with a plurality of positions in the genome, and an associated probability belonging to that position.

Counting Method to Determine Chromosome Copy Number 10 In one aspect, the invention features methods of testing for an abnormal distribution of a fetal chromosome by comparing the number of sequence tags that align to different chromosomes (see, e.g., U.S. Pat. No. 8,296,076, filed Apr. 20, 2012, which is hereby incorporated by reference in 15 its entirety). As is known in the art, the term "sequence tag" refers to a relatively short (e.g., 15-100) nucleic acid sequence that can be used to identify a certain larger sequence, e.g., be mapped to a chromosome or genomic region or gene. In some embodiments, the method involves 20 (i) contacting a sample that includes a mixture of maternal and fetal DNA with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 25 100,000 different target loci to produce a reaction mixture; wherein the target loci are from a plurality of different chromosomes; and wherein the plurality of different chromosomes comprise at least one first chromosome suspected of having an abnormal distribution in the sample and at least 30 one second chromosome presumed to be normally distributed in the sample; (ii) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products; (iii) sequencing the amplified products to obtain a plurality of sequence tags aligning to the target loci; wherein 3 the sequence tags are of sufficient length to be assigned to a specific target locus; (iv) assigning on a computer the plurality of sequence tags to their corresponding target loci; (v) determining on a computer a number of sequence tags aligning to the target loci of the first chromosome and a 40 number of sequence tags aligning to the target loci of the second chromosome; and (vi) comparing the numbers from step (v) to determine the presence or absence of an abnormal distribution of the first chromosome.

In one aspect, the invention provides methods for detect- 45 ing the presence or absence of a fetal aneuploidy by comparing the relative frequency of target amplicons between chromosomes (see, e.g., PCT Publ. No. WO 2012/103031, filed Jan. 23, 2012, which is hereby incorporated by reference in its entirety). In some embodiments, the method 50 involves (i) contacting a sample that includes a mixture of maternal and fetal DNA with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 55 50,000; 75,000; or 100,000 different non-polymorphic target loci to produce a reaction mixture; wherein the target loci are from a plurality of different chromosomes; (ii) subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that includes target amplicons; 60 (iii) quantifying on a computer a relative frequency of the target amplicons from the first and second chromosomes of interest; (iv) comparing on a computer the relative frequency of the target amplicons from the first and second chromosomes of interest; and (v) identifying the presence or 65 absence of an aneuploidy based on the compared relative frequencies of the first and second chromosome of interest.

188

In some embodiments, the first chromosome is a chromosome suspected of being euploid. In some embodiments, the second chromosome is a chromosome suspected of being aneuploidy

Combining Methods of Prenatal Diagnosis

There are many methods that may be used for prenatal diagnosis or prenatal screening of aneuploidy or other genetic defects. Described elsewhere in this document, and in U.S. Utility application Ser. No. 11/603,406, filed Nov. 28, 2006; U.S. Utility application Ser. No. 12/076,348, filed Mar. 17, 2008, and PCT Application Serial No. PCT/S09/ 52730 is one such method that uses the genetic data of related individuals to increase the accuracy with which genetic data of a target individual, such as a fetus, is known, or estimated. Other methods used for prenatal diagnosis involve measuring the levels of certain hormones in maternal blood, where those hormones are correlated with various genetic abnormalities. An example of this is called the triple test, a test wherein the levels of several (commonly two, three, four or five) different hormones are measured in maternal blood. In a case where multiple methods are used to determine the likelihood of a given outcome, where none of the methods are definitive in and of themselves, it is possible to combine the information given by those methods to make a prediction that is more accurate than any of the individual methods. In the triple test, combining the information given by the three different hormones can result in a prediction of genetic abnormalities that is more accurate than the individual hormone levels may predict.

Disclosed herein is a method for making more accurate predictions about the genetic state of a fetus, specifically the possibility of genetic abnormalities in a fetus that comprises combining predictions of genetic abnormalities in a fetus where those predictions were made using a variety of methods. A "more accurate" method may refer to a method for diagnosing an abnormality that has a lower false negative rate at a given false positive rate. In a favored embodiment of the present disclosure, one or more of the predictions are made based on the genetic data known about the fetus, where the genetic knowledge was determined using the PARENTAL SUPPORT<sup>TM</sup> method, that is, using genetic data of individual related to the fetus to determine the genetic data of the fetus with greater accuracy. In some embodiments the genetic data may include ploidy states of the fetus. In some embodiments, the genetic data may refer to a set of allele calls on the genome of the fetus. In some embodiments some of the predictions may have been made using the triple test. In some embodiments, some of the predictions may have been made using measurements of other hormone levels in maternal blood. In some embodiments, predictions made by methods considered diagnoses may be combined with predictions made by methods considered screening. In some embodiments, the method involves measuring maternal blood levels of alpha-fetoprotein (AFP). In some embodiments, the method involves measuring maternal blood levels of unconjugated estriol (UE<sub>3</sub>). In some embodiments, the method involves measuring maternal blood levels of beta human chorionic gonadotropin (beta-hCG). In some embodiments, the method involves measuring maternal blood levels of invasive trophoblast antigen (ITA). In some embodiments, the method involves measuring maternal blood levels of inhibin. In some embodiments, the method involves measuring maternal blood levels of pregnancy-associated plasma protein A (PAPP-A). In some embodiments, the method involves measuring maternal blood levels of other hormones or maternal serum markers. In some embodiments, some of the

# US 11,519,035 B2

189

predictions may have been made using other methods. In some embodiments, some of the predictions may have been made using a fully integrated test such as one that combines ultrasound and blood test at around 12 weeks of pregnancy and a second blood test at around 16 weeks. In some embodiments, the method involves measuring the fetal nuchal translucency (NT). In some embodiments, the method involves using the measured levels of the aforementioned hormones for making predictions. In some embodiments the method involves a combination of the aforementioned methods.

There are many ways to combine the predictions, for example, one could convert the hormone measurements into a multiple of the median (MoM) and then into likelihood ratios (LR). Similarly, other measurements could be transformed into LRs using the mixture model of NT distributions. The LRs for NT and the biochemical markers could be multiplied by the age and gestation-related risk to derive the risk for various conditions, such as trisomy 21. Detection rates (DRs) and false-positive rates (FPRs) could be calculated by taking the proportions with risks above a given risk threshold.

In an embodiment, a method to call the ploidy state involves combining the relative probabilities of each of the 25 ploidy hypotheses determined using the joint distribution model and the allele count probabilities with relative probabilities of each of the ploidy hypotheses that are calculated using statistical techniques taken from other methods that determine a risk score for a fetus being trisomic, including 30 but not limited to: a read count analysis, comparing heterozygosity rates, a statistic that is only available when parental genetic information is used, the probability of normalized genotype signals for certain parent contexts, a statistic that is calculated using an estimated fetal fraction of 35 the first sample or the prepared sample, and combinations thereof.

Another method could involve a situation with four measured hormone levels, where the probability distribution around those hormones is known:  $p(x, x_2, x_3, x_4|e)$  for the 40 euploid case and  $p(x_1, x_2, x_3, x_4|a)$  for the aneuploid case. Then one could measure the probability distribution for the DNA measurements, g(y|e) and g(y|a) for the euploid and aneuploid cases respectively. Assuming they are independent given the assumption of euploid/aneuploid, one could 45 combine as  $p(x_1, x_2, x_3, x_4|a)g(y|a)$  and  $p(x_1, x_2, x_3, x_4|e)$  g(y|e) and then multiply each by the prior p(a) and p(e) given the maternal age. One could then choose the one that is highest.

In an embodiment, it is possible to evoke central limit 50 theorem to assume distribution on g(y|a or e) is Gaussian, and measure mean and standard deviation by looking at multiple samples. In another embodiment, one could assume they are not independent given the outcome and collect enough samples to estimate the joint distribution  $p(x_1, x_2, x_3, 55 x_4|a \text{ or e})$ .

In an embodiment, the ploidy state for the target individual is determined to be the ploidy state that is associated with the hypothesis whose probability is the greatest. In some cases, one hypothesis will have a normalized, combined probability greater than 90%. Each hypothesis is associated with one, or a set of, ploidy states, and the ploidy state associated with the hypothesis whose normalized, combined probability is greater than 90%, or some other threshold value, such as 50%, 80%, 95%, 98%, 99%, or 65 99.9%, may be chosen as the threshold required for a hypothesis to be called as the determined ploidy state.

190

DNA from Children from Previous Pregnancies in Maternal Blood

One difficulty to non-invasive prenatal diagnosis is differentiating fetal cells from the current pregnancy from fetal cells from previous pregnancies. Some believe that genetic matter from prior pregnancies will go away after some time, but conclusive evidence has not been shown. In an embodiment of the present disclosure, it is possible to determine fetal DNA present in the maternal blood of paternal origin (that is, DNA that the fetus inherited from the father) using the PARENTAL SUPPORT<sup>TM</sup> (PS) method, and the knowledge of the paternal genome. This method may utilize phased parental genetic information. It is possible to phase the parental genotype from unphased genotypic information using grandparental genetic data (such as measured genetic data from a sperm from the grandfather), or genetic data from other born children, or a sample of a miscarriage. One could also phase unphased genetic information by way of a HapMap-based phasing, or a haplotyping of paternal cells. Successful haplotyping has been demonstrated by arresting cells at phase of mitosis when chromosomes are tight bundles and using microfluidics to put separate chromosomes in separate wells. In another embodiment it is possible to use the phased parental haplotypic data to detect the presence of more than one homolog from the father, implying that the genetic material from more than one child is present in the blood. By focusing on chromosomes that are expected to be euploid in a fetus, one could rule out the possibility that the fetus was afflicted with a trisomy. Also, it is possible to determine if the fetal DNA is not from the current father, in which case one could use other methods such as the triple test to predict genetic abnormalities.

There may be other sources of fetal genetic material available via methods other than a blood draw. In the case of the fetal genetic material available in maternal blood, there are two main categories: (1) whole fetal cells, for example, nucleated fetal red blood cells or erythroblats, and (2) free floating fetal DNA. In the case of whole fetal cells, there is some evidence that fetal cells can persist in maternal blood for an extended period of time such that it is possible to isolate a cell from a pregnant woman that contains the DNA from a child or fetus from a prior pregnancy. There is also evidence that the free floating fetal DNA is cleared from the system in a matter of weeks. One challenge is how to determine the identity of the individual whose genetic material is contained in the cell, namely to ensure that the measured genetic material is not from a fetus from a prior pregnancy. In an embodiment of the present disclosure, the knowledge of the maternal genetic material can be used to ensure that the genetic material in question is not maternal genetic material. There are a number of methods to accomplish this end, including informatics based methods such as PARENTAL SUPPORT<sup>TM</sup>, as described in this document or any of the patents referenced in this document.

In an embodiment of the present disclosure, the blood drawn from the pregnant mother may be separated into a fraction comprising free floating fetal DNA, and a fraction comprising nucleated red blood cells. The free floating DNA may optionally be enriched, and the genotypic information of the DNA may be measured. From the measured genotypic information from the free floating DNA, the knowledge of the maternal genotype may be used to determine aspects of the fetal genotype. These aspects may refer to ploidy state, and/or a set of allele identities. Then, individual nucleated red blood cells may be genotyped using methods described elsewhere in this document, and other referent patents, especially those mentioned in the first section of this docu-

# US 11,519,035 B2

19

ment. The knowledge of the maternal genome would allow one to determine whether or not any given single blood cell is genetically maternal. And the aspects of the fetal genotype that were determined as described above would allow one to determine if the single blood cell is genetically derived from the fetus that is currently gestating. In essence, this aspect of the present disclosure allows one to use the genetic knowledge of the mother, and possibly the genetic information from other related individuals, such as the father, along with the measured genetic information from the free floating 10 DNA found in maternal blood to determine whether an isolated nucleated cell found in maternal blood is either (a) genetically maternal, (b) genetically from the fetus currently gestating, or (c) genetically from a fetus from a prior pregnancy.

Prenatal Sex Chromosome Aneuploidy Determination

In methods known in the art, people attempting to determine the sex of a gestating fetus from the blood of the mother have used the fact that fetal free floating DNA (fffDNA) is present in the plasma of the mother. If one is 20 able to detect Y-specific loci in the maternal plasma, this implies that the gestating fetus is a male. However, the lack of detection of Y-specific loci in the plasma does not always guarantee that the gestating fetus is a female when using methods known in the art, as in some cases the amount of 25 fffDNA is too low to ensure that the Y-specific loci would be detected in the case of a male fetus.

Presented here is a novel method that does not require the measurement of Y-specific nucleic acids, that is, DNA that is from loci that are exclusively paternally derived. The 30 Parental Support method, disclosed previously, uses crossover frequency data, parental genotypic data, and informatics techniques, to determine the ploidy state of a gestating fetus. The sex of a fetus is simply the ploidy state of the fetus at the sex chromosomes. A child that is XX is female, and 35 XY is male. The method described herein is also able to determine the ploidy state of the fetus. Note that sexing is effectively synonymous with ploidy determination of the sex chromosomes; in the case of sexing, an assumption is often made that the child is euploid, therefore there are fewer 40 possible hypotheses.

The method disclosed herein involves looking at loci that are common to both the X and Y chromosome to create a baseline in terms of expected amount of fetal DNA present for a fetus. Then, those regions that are specific only to the 45 X chromosome can be interrogated to determine if the fetus is female or male. In the case of a male, we expect to see less fetal DNA from loci that are specific to the X chromosome than from loci that are specific to both the X and the Y. In contrast, in female fetuses, we expect the amount of DNA for each of these groups to be the same. The DNA in question can be measured by any technique that can quantitate the amount of DNA present on a sample, for example, qPCR, SNP arrays, genotyping arrays, or sequencing. For DNA that is exclusively from an individual we would expect 55 to see the following:

	DNA specific to X	DNA specific to X and Y	DNA specific to Y
Male (XY)	A	2A	A
Female (XX)	2A	2A	0

In the case of DNA from a fetus that is mixed with DNA from the mother, and where the fraction of fetal DNA in the

192

mixture is F, and where the fraction of maternal DNA in the mixture is M, such that F+M=100%, we would expect to see the following:

`					
_		DNA specific to X	DNA specific to X and Y	DNA specific to Y	
	Male fetus (XY)	M + ½ F	M + F	½ F	
0	Female fetus (XX)	M + F	M + F	0	

In the case where F and M are known, the expected ratios can be computed, and the observed data can be compared to the expected data. In the case where M and F are not known, a threshold can be selected based on historical data. In both cases, the measured amount of DNA at loci specific to both X and Y can be used as a baseline, and the test for the sex of the fetus can be based on the amount of DNA observed on loci specific to only the X chromosome. If that amount is lower than the baseline by an amount roughly equal to ½ F, or by an amount that causes it to fall below a predefined threshold, the fetus is determined to be male, and if that amount is about equal to the baseline, or if is not lower by an amount that causes it to fall below a predefined threshold, the fetus is determined to be female.

In another embodiment, one can look only at those loci that are common to both the X and the Y chromosomes, often termed the Z chromosome. A subset of the loci on the Z chromosome are typically always A on the X chromosome, and B on the Y chromosome. If SNPs from the Z chromosome are found to have the B genotype, then the fetus is called a male; if the SNPs from the Z chromosome are found to only have A genotype, then the fetus is called a female. In another embodiment, one can look at the loci that are found only on the X chromosome. Contexts such as AAIB are particularly informative as the presence of a B indicates that the fetus has an X chromosome from the father. Contexts such as ABIB are also informative, as we expect to see B present only half as often in the case of a female fetus as compared to a male fetus. In another embodiment, one can look at the SNPs on the Z chromosome where both A and B alleles are present on both the X and the Y chromosome, and where the it is known which SNPs are from the paternal Y chromosome, and which are from the paternal X chromosome.

In an embodiment, it is possible to amplify single nucleotide positions known to varying between the homologous non-recombining (HNR) region shared by chromosome Y and chromosome X. The sequence within this HNR region is largely identical between the X and Y chromosomes. Within this identical region are single nucleotide positions that, while invariant among X chromosomes and among Y chromosomes in the population, are different between the X and Y chromosomes. Each PCR assay could amplify a sequence from loci that are present on both the X and Y chromosomes. Within each amplified sequence would be a single base that can be detected using sequencing or some other method (see, for example, U.S. Publication No. 2011/60 0178719, filed Feb. 3, 2011, which is hereby incorporated by reference in its entirety).

In an embodiment, the sex of the fetus could be determined from the fetal free floating DNA found in maternal plasma, the method comprising some or all of the following steps: 1) Design PCR (either regular or mini-PCR, plus multiplexing if desired) primers amplify X/Y variant single nucleotide positions within HNR region, 2) obtain maternal

# US 11,519,035 B2

193

plasma, 3) PCR Amplify targets from maternal plasma using HNR X/Y PCR assays, 4) sequence the amplicons, 5) Examine sequence data for presence of Y-allele within one or more of the amplified sequences. The presence of one or more would indicate a male fetus. Absence of all Y-alleles 5 from all amplicons indicates a female fetus.

In an embodiment, one could use targeted sequencing to measure the DNA in the maternal plasma and/or the parental genotypes. In an embodiment, one could ignore all sequences that clearly originate from paternally sourced 10 DNA. For example, in the context AA|AB, one could count the number of A sequences and ignore all the B sequences. In order to determine a heterozygosity rate for the above algorithm, one could compare the number of observed A sequences to the expected number of total sequences for the 15 given probe. There are many ways one could calculate an expected number of sequences for each probe on a per sample basis. In an embodiment, it is possible to use historical data to determine what fraction of all sequence reads belongs to each specific probe and then use this 20 empirical fraction, combined with the total number of sequence reads, to estimate the number of sequences at each probe. Another approach could be to target some known homozygous alleles and then use historical data to relate the number of reads at each probe with the number of reads at 25 the known homozygous alleles. For each sample, one could then measure the number of reads at the homozygous alleles and then use this measurement, along with the empirically derived relationships, to estimate the number of sequence reads at each probe.

In some embodiments, it is possible to determine the sex of the fetus by combining the predictions made by a plurality of methods. In some embodiments the plurality of methods are taken from methods described in this disclosure. In some embodiments, at least one of the plurality of methods are 35 taken from methods described in this disclosure.

In some embodiments the method described herein can be used to determine the ploidy state of the gestating fetus. In an embodiment, the ploidy calling method uses loci that are specific to the X chromosome, or common to both the X and 40 Y chromosome, but does not make use of any Y-specific loci. In an embodiment, the ploidy calling method uses one or more of the following: loci that are specific to the X chromosome, loci that are common to both the X and Y chromosome, and loci that are specific to the Y chromosome. 45 In an embodiment, where the ratios of sex chromosomes are similar, for example 45,X (Turner Syndrome), 46,XX (normal female) and 47,XXX (trisomy X), the differentiation can be accomplished by comparing the allele distributions to expected allele distributions according to the various 50 hypotheses. In another embodiment, this can be accomplished by comparing the relative number of sequence reads for the sex chromosomes to one or a plurality of reference chromosomes that are assumed to be euploid. Also note that these methods can be expanded to include aneuploid cases. 55 Single Gene Disease Screening

In an embodiment, a method for determining the ploidy state of the fetus may be extended to enable simultaneous testing for single gene disorders. Single-gene disease diagnosis leverages the same targeted approach used for aneuploidy testing, and requires additional specific targets. In an embodiment, the single gene NPD diagnosis is through linkage analysis. In many cases, direct testing of the cfDNA sample is not reliable, as the presence of maternal DNA makes it virtually impossible to determine if the fetus has inherited the mother's mutation. Detection of a unique paternally-derived allele is less challenging, but is only fully

194

informative if the disease is dominant and carried by the father, limiting the utility of the approach. In an embodiment, the method involves PCR or related amplification approaches.

In some embodiments, the method involves phasing the abnormal allele with surrounding very tightly linked SNPs in the parents using information from first-degree relatives. Then Parental Support may be run on the targeted sequencing data obtained from these SNPs to determine which homologs, normal or abnormal, were inherited by the fetus from both parents. As long as the SNPs are sufficiently linked, the inheritance of the genotype of the fetus can be determined very reliably. In some embodiments, the method comprises (a) adding a set of SNP loci to densely flank a specified set of common diseases to our multiplex pool for aneuploidy testing; (b) reliably phasing the alleles from these added SNPs with the normal and abnormal alleles based on genetic data from various relatives; and (c) reconstructing the fetal haplotype, or set of phased SNP alleles on the inherited maternal and paternal homologs in the region surrounding the disease locus to determine fetal genotype. In some embodiments additional probes that are closely linked to a disease linked locus are added to the set of polymorphic locus being used for aneuploidy testing.

Reconstructing fetal diplotype is challenging because the sample is a mixture of maternal and fetal DNA. In some embodiments, the method incorporates relative information to phase the SNPs and disease alleles, then take into account physical distance of the SNPs and recombination data from location specific recombination likelihoods and the data observed from the genetic measurements of the maternal plasma to obtain the most likely genotype of the fetus.

In an embodiment, a number of additional probes per disease linked locus are included in the set of targeted polymorphic loci; the number of additional probes per disease linked locus may be between 4 and 10, between 11 and 20, between 21 and 40, between 41 and 60, between 61 and 80, or combinations thereof.

Phasing the diploid data from the parents can be challenging, and there are a number of ways this can be accomplished. Some are discussed in this disclosure, others are described in greater detail in other disclosures (see, e.g., PCT Publ. No. WO2009105531, filed Feb. 9, 2009, and PCT Publ. No. WO2010017214, filed Aug. 4, 2009, which are each hereby incorporated by reference in its entirety). In one embodiment, a parent can be phased by inference by measuring tissue from the parent that is haploid, for example by measuring one or more sperm or eggs. In one embodiment the parent can be phased by inference using the measured genotypic data of a first degree relative such as the parent's parent(s) or siblings. In one embodiment, the parent can be phased by dilution where the DNA is diluted, in one or a plurality of wells, to the point where there is expected to be no more than approximately one copy of each haplotype in each well, and then measuring the DNA in the one or more wells. In one embodiment, the parent genotype can be phased by using computer programs that use population based haplotype frequencies to infer the most likely phase. In one embodiment, the parent can be phased if the phased haplotypic data is known for the other parent, along with the unphased genetic data of one or more genetic offspring of the parents. In some embodiments, the genetic offspring of the parents may be one or more embryos, fetuses, and/or born children. Some of these methods and other methods for phasing one or both parents are disclosed in greater detail in, e.g., U.S. Publ. No. 2011/0033862, filed Aug. 19, 2010; U.S. Publ. No. 2011/0178719, filed Feb. 3, 2011; U.S. Publ. No.

# US 11,519,035 B2

19

2007/0184467, filed Nov. 22, 2006; U.S. Publ. No. 2008/0243398, filed Mar. 17, 2008, which are each hereby incorporated by reference in its entirety.

Fetal Genome Reconstruction

In one aspect, the invention features methods for determining a haplotype of a fetus. In various embodiments, this method allows one to determine which polymorphic loci (such as SNPs) were inherited by the fetus and to reconstruct which homologs (including recombination events) are present in the fetus (and thereby interpolate the sequence 10 between the polymorphic loci). If desired, essentially the entire genome of the fetus can be reconstructed. If there is some remaining ambiguity in the genome of the fetus (such as in intervals with a crossover), this ambiguity can be minimized if desired by analyzing additional polymorphic 15 loci. In various embodiments, the polymorphic loci are chosen to cover one or more of the chromosomes at a density to reduce any ambiguity to a desired level. This method has important applications for the detection of polymorphisms or other mutations of interest in a fetus since it enables their 20 detection based on linkage (such as the presence of linked polymorphic loci in the fetal genome) rather than by directing detecting the polymorphism or other mutation of interest in the fetal genome. For example, if a parent is a carrier for a mutation associated with cystic fibrosis (CF), a nucleic 25 acid sample that includes maternal DNA from the mother of the fetus and fetal DNA from the fetus can be analyzed to determine whether the fetal DNA include the haplotype containing the CF mutation. In particular, polymorphic loci can be analyzed to determine whether the fetal DNA 30 includes the haplotype containing the CF mutation without having to detect the CF mutation itself in the fetal DNA. This is useful in screening for one or more mutations, such as disease-linked mutations, without having to directly detect the mutations.

In some embodiments, the method involves determining a parental haplotype (e.g., a haplotype of the mother or father of the fetus). In some embodiments, this determination is made without using data from a relative of the mother or father. In some embodiments, a parental haplotype is 40 determined using a dilution approach followed by SNP genotyping or sequencing as described herein and elsewhere (see, e.g., U.S. Publ. No. 2011/0033862, filed Aug. 19, 2010, which is hereby incorporated by reference in its entirety). Because the DNA is diluted, it is unlikely that more than one 45 haplotype is in the same fraction (or tube). Thus, there may be effectively a single molecule of DNA in the tube, which allows the haplotype on a single DNA molecule to be determined. In some embodiments, the method includes dividing a DNA sample into a plurality of fractions such that 50 at least one of the fractions includes one chromosome or one chromosome segment from a pair of chromosomes, and genotyping (e.g., determining the presence of two or more polymorphic loci) the DNA sample in at least one of the fractions, thereby determining a parental haplotype. In some 55 embodiments, the genotyping involves sequencing (such as shotgun sequencing). In some embodiments, the genotyping involves use of a SNP array to detect polymorphic loci, such as at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 60 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci. In some embodiments, the genotyping involves the use of multiplex PCR. In some embodiments, the method involves contacting the sample in a fraction with a library of primers that simultaneously hybrid- 65 ize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000;

196

27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci (such as SNPs) to produce a reaction mixture; and subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that are measured with a high throughput sequencer to produce sequencing data.

In some embodiments, a haplotype of the mother is determined by any of the methods described herein using data from a relative of the mother. In some embodiments, a haplotype of the father is determined by any of the methods described herein using data from a relative of the father. In some embodiments, a haplotype is determined for both the father and the mother. In some embodiments, a SNP array is used to determine the presence of at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000;  $19,000;\ 20,000;\ 25,000;\ 27,000;\ 28,000;\ 30,000;\ 40,000;$ 50,000; 75,000; or 100,000 different polymorphic loci in a DNA sample from the mother (or father) and a relative of the mother (or father). In some embodiments, the method involves contacting a DNA sample from the mother (or father) and/or a relative of the mother (or father) with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci (such as SNPs) to produce a reaction mixture; and subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that are measured with a high throughput sequencer to produce sequencing data. The parental haplotype may be determined based on the SNP array or sequencing data. In some embodiments, parental data may be phased by methods described or referred to elsewhere in this document.

This parental haplotype data can be used to determine if the fetus inherited the parental haplotype. In some embodiments, a nucleic acid sample that includes maternal DNA from the mother of the fetus and fetal DNA from the fetus is analyzed using a SNP array to detect at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci.

In some embodiments, a nucleic acid sample that includes maternal DNA from the mother of the fetus and fetal DNA from the fetus is analyzed by contacting the sample with a library of primers that simultaneously hybridize to at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci (such as SNPs) to produce a reaction mixture. In some embodiments, the reaction mixture is subjected to primer extension reaction conditions to produce amplified products. In some embodiments, the amplified products are measured with a high throughput sequencer to produce sequencing data. In various embodiments, the SNP array or sequencing data is used to determine a parental haplotype by using data about the probability of chromosomes crossing over at different locations in a chromosome (such as by using recombination data such as may be found in the HapMap database to create a recombination risk score for any interval) to model dependence between polymorphic alleles on the chromosome. In some embodiments, allele counts at the polymorphic loci are calculated on a computer based on the sequencing data. In some embodiments, a plurality of ploidy hypotheses each pertaining to a different possible ploidy state of the chromosome are created on a computer; a model (such as a joint distribution model) for

# US 11,519,035 B2

197

the expected allele counts at the polymorphic loci on the chromosome is built on a computer for each ploidy hypothesis; a relative probability of each of the ploidy hypotheses is determined on a computer using the joint distribution model and the allele counts; and the ploidy state of the fetus is called by selecting the ploidy state corresponding to the hypothesis with the greatest probability. In some embodiments, building a joint distribution model for allele counts and the step of determining the relative probability of each hypothesis are done using a method that does not require the use of a reference chromosome.

In some embodiments, a fetal haplotype is determined for one or more chromosomes taken from the group consisting of chromosomes 13, 18, 21, X, and Y. In some embodiments, a fetal haplotype is determined for all of the fetal chromosomes. In various embodiments, the method determines essentially the entire genome of the fetus. In some embodiments, the haplotype is determined for at least 30, 40, 50, 60, 70, 80, 90, or 95% of the genome of the fetus.

In some embodiments, the haplotype determination of the fetus includes information about which allele is present for at least 25; 50; 75; 100; 300; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 15,000; 19,000; 20,000; 25,000; 27,000; 28,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different 25 polymorphic loci.

Compositions of DNA

When performing an informatics analysis on sequencing data measured on a mixture of fetal and maternal blood to determine genomic information pertaining to the fetus, for 30 example the ploidy state of the fetus, it may be advantageous to measure the allele distributions at a set of alleles. Unfortunately, in many cases, such as when attempting to determine the ploidy state of a fetus from the DNA mixture found in the plasma of a maternal blood sample, the amount of 3 DNA available is not sufficient to directly measure the allele distributions with good fidelity in the mixture. In these cases, amplification of the DNA mixture will provide sufficient numbers of DNA molecules that the desired allele distributions may be measured with good fidelity. However, 40 current methods of amplification typically used in the amplification of DNA for sequencing are often very biased, meaning that they do not amplify both alleles at a polymorphic locus by the same amount. A biased amplification can result in allele distributions that are quite different from the 45 allele distributions in the original mixture. For most purposes, highly accurate measurements of the relative amounts of alleles present at polymorphic loci are not needed. In contrast, in an embodiment of the present disclosure, amplification or enrichment methods that specifically enrich polymorphic alleles and preserve allelic ratios is advantageous.

A number of methods are described herein that may be used to preferentially enrich a sample of DNA at a plurality of loci in a way that minimizes allelic bias. Some examples are using circularizing probes to target a plurality of loci 55 where the 3' ends and 5' ends of the pre-circularized probe are designed to hybridize to bases that are one or a few positions away from the polymorphic sites of the targeted allele. Another is to use PCR probes where the 3' end PCR probe is designed to hybridize to bases that are one or a few positions away from the polymorphic sites of the targeted allele. Another is to use a split and pool approach to create mixtures of DNA where the preferentially enriched loci are enriched with low allelic bias without the drawbacks of direct multiplexing. Another is to use a hybrid capture 65 approach where the capture probes are designed such that the region of the capture probe that is designed to hybridize

198

to the DNA flanking the polymorphic site of the target is separated from the polymorphic site by one or a small number of bases.

In the case where measured allele distributions at a set of polymorphic loci are used to determine the ploidy state of an individual, it is desirable to preserve the relative amounts of alleles in a sample of DNA as it is prepared for genetic measurements. This preparation may involve WGA amplification, targeted amplification, selective enrichment techniques, hybrid capture techniques, circularizing probes or other methods meant to amplify the amount of DNA and/or selectively enhance the presence of molecules of DNA that correspond to certain alleles.

In some embodiments of the present disclosure, there is a set of DNA probes designed to target loci where the loci have maximal minor allele frequencies. In some embodiments of the present disclosure, there is a set of probes that are designed to target where the loci have the maximum likelihood of the fetus having a highly informative SNP at 20 those loci. In some embodiments of the present disclosure, there is a set of probes that are designed to target loci where the probes are optimized for a given population subgroup. In some embodiments of the present disclosure, there is a set of probes that are designed to target loci where the probes are optimized for a given mix of population subgroups. In some embodiments of the present disclosure, there is a set of probes that are designed to target loci where the probes are optimized for a given pair of parents which are from different population subgroups that have different minor allele frequency profiles. In some embodiments of the present disclosure, there is a circularized strand of DNA that comprises at least one base pair that annealed to a piece of DNA that is of fetal origin. In some embodiments of the present disclosure, there is a circularized strand of DNA that comprises at least one base pair that annealed to a piece of DNA that is of placental origin. In some embodiments of the present disclosure, there is a circularized strand of DNA that circularized while at least some of the nucleotides were annealed to DNA that was of fetal origin. In some embodiments of the present disclosure, there is a circularized strand of DNA that circularized while at least some of the nucleotides were annealed to DNA that was of placental origin. In some embodiments of the present disclosure, there is a set of probes wherein some of the probes target single tandem repeats, and some of the probes target single nucleotide polymorphisms. In some embodiments, the loci are selected for the purpose of non-invasive prenatal diagnosis. In some embodiments, the probes are used for the purpose of noninvasive prenatal diagnosis. In some embodiments, the loci are targeted using a method that could include circularizing probes, MIPs, capture by hybridization probes, probes on a SNP array, or combinations thereof. In some embodiments, the probes are used as circularizing probes, MIPs, capture by hybridization probes, probes on a SNP array, or combinations thereof. In some embodiments, the loci are sequenced for the purpose of non-invasive prenatal diagnosis.

In the case where the relative informativeness of a sequence is greater when combined with relevant parent contexts, it follows that maximizing the number of sequence reads that contain a SNP for which the parental context is known may maximize the informativeness of the set of sequencing reads on the mixed sample. In an embodiment, the number of sequence reads that contain a SNP for which the parent contexts are known may be enhanced by using qPCR to preferentially amplify specific sequences. In an embodiment, the number of sequence reads that contain a SNP for which the parent contexts are known may be

# US 11,519,035 B2

199

enhanced by using circularizing probes (for example, MIPs) to preferentially amplify specific sequences. In an embodiment, the number of sequence reads that contain a SNP for which the parent contexts are known may be enhanced by using a capture by hybridization method (for example 5 SURESELECT) to preferentially amplify specific sequences. Different methods may be used to enhance the number of sequence reads that contain a SNP for which the parent contexts are known. In an embodiment, the targeting may be accomplished by extension ligation, ligation without 10 extension, capture by hybridization, or PCR.

In a sample of fragmented genomic DNA, a fraction of the DNA sequences map uniquely to individual chromosomes; other DNA sequences may be found on different chromosomes. Note that DNA found in plasma, whether maternal or 15 fetal in origin is typically fragmented, often at lengths under 500 bp. In a typical genomic sample, roughly 3.3% of the mappable sequences will map to chromosome 13; 2.2% of the mappable sequences will map to chromosome 18; 1.35% of the mappable sequences will map to chromosome 21; 20 4.5% of the mappable sequences will map to chromosome X in a female; 2.25% of the mappable sequences will map to chromosome X (in a male); and 0.73% of the mappable sequences will map to chromosome Y (in a male). These are the chromosomes that are most likely to be an uploid in a 25 fetus. Also, among short sequences, approximately 1 in 20 sequences will contain a SNP, using the SNPs contained on dbSNP. The proportion may well be higher given that there may be many SNPs that have not been discovered.

In an embodiment of the present disclosure, targeting 30 methods may be used to enhance the fraction of DNA in a sample of DNA that map to a given chromosome such that the fraction significantly exceeds the percentages listed above that are typical for genomic samples. In an embodiment of the present disclosure, targeting methods may be used to enhance the fraction of DNA in a sample of DNA such that the percentage of sequences that contain a SNP are significantly greater than what may be found in typical for genomic samples. In an embodiment of the present disclosure, targeting methods may be used to target DNA from a 40 chromosome or from a set of SNPs in a mixture of maternal and fetal DNA for the purposes of prenatal diagnosis.

Note that a method has been reported (U.S. Pat. No. 7,888,017) for determining fetal aneuploidy by counting the number of reads that map to a suspect chromosome and 45 comparing it to the number of reads that map to a reference chromosome, and using the assumption that an overabundance of reads on the suspect chromosome corresponds to a triploidy in the fetus at that chromosome. Those methods for prenatal diagnosis would not make use of targeting of any 50 sort, nor do they describe the use of targeting for prenatal diagnosis.

By making use of targeting approaches in sequencing the mixed sample, it may be possible to achieve a certain level of accuracy with fewer sequence reads. The accuracy may 55 refer to sensitivity, it may refer to specificity, or it may refer to some combination thereof. The desired level of accuracy may be between 90% and 95%; it may be between 95% and 98%; it may be between 98% and 99%; it may be between 99% and 99.5%; it may be between 99.9% and 99.99%; it may be between 99.99% and 99.99%, it may be between 99.99% and 100%. Levels of accuracy above 95% may be referred to as high accuracy.

There are a number of published methods in the prior art 65 that demonstrate how one may determine the ploidy state of a fetus from a mixed sample of maternal and fetal DNA, for

200

example: G. J. W. Liao et al. Clinical Chemistry 2011; 57(1) pp. 92-101. These methods focus on thousands of locations along each chromosome. The number of locations along a chromosome that may be targeted while still resulting in a high accuracy ploidy determination on a fetus, for a given number of sequence reads, from a mixed sample of DNA is unexpectedly low. In an embodiment of the present disclosure, an accurate ploidy determination may be made by using targeted sequencing, using any method of targeting, for example qPCR, ligand mediated PCR, other PCR methods, capture by hybridization, or circularizing probes, wherein the number of loci along a chromosome that need to be targeted may be between 5,000 and 2,000 loci; it may be between 2,000 and 1,000 loci; it may be between 1,000 and 500 loci; it may be between 500 and 300 loci; it may be between 300 and 200 loci; it may be between 200 and 150 loci; it may be between 150 and 100 loci; it may be between 100 and 50 loci; it may be between 50 and 20 loci; it may be between 20 and 10 loci. Optimally, it may be between 100 and 500 loci. The high level of accuracy may be achieved by targeting a small number of loci and executing an unexpectedly small number of sequence reads. The number of reads may be between 100 million and 50 million reads; the number of reads may be between 50 million and 20 million reads; the number of reads may be between 20 million and  $10 \ million$  reads; the number of reads may be between 10million and 5 million reads; the number of reads may be between 5 million and 2 million reads; the number of reads may be between 2 million and 1 million; the number of reads may be between 1 million and 500,000; the number of reads may be between 500,000 and 200,000; the number of reads may be between 200,000 and 100,000; the number of reads may be between 100,000 and 50,000; the number of reads may be between 50,000 and 20,000; the number of reads may be between 20,000 and 10,000; the number of reads may be below 10,000. Fewer number of read are necessary for larger amounts of input DNA.

In some embodiments, there is a composition comprising a mixture of DNA of fetal origin, and DNA of maternal origin, wherein the percent of sequences that uniquely map to chromosome 13 is greater than 4%, greater than 5%, greater than 6%, greater than 7%, greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, greater than 20%, greater than 25%, or greater than 30%. In some embodiments of the present disclosure, there is a composition comprising a mixture of DNA of fetal origin, and DNA of maternal origin, wherein the percent of sequences that uniquely map to chromosome 18 is greater than 3%, greater than 4%, greater than 5%, greater than 6%, greater than 7%, greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, greater than 20%, greater than 25%, or greater than 30%. In some embodiments of the present disclosure, there is a composition comprising a mixture of DNA of fetal origin, and DNA of maternal origin, wherein the percent of sequences that uniquely map to chromosome 21 is greater than 2%, greater than 3%, greater than 4%, greater than 5%, greater than 6%, greater than 7%, greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, greater than 20%, greater than 25%, or greater than 30%. In some embodiments of the present disclosure, there is a composition comprising a mixture of DNA of fetal origin, and DNA of maternal origin, wherein the percent of sequences that uniquely map to chromosome X is greater than 6%, greater than 7%, greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, greater than 20%, greater than 25%, or greater than 30%. In some embodiCase: 24-1324 Document: 42-1 Page: 223 Filed: 03/18/2024

# US 11,519,035 B2

ments of the present disclosure, there is a composition comprising a mixture of DNA of fetal origin, and DNA of maternal origin, wherein the percent of sequences that uniquely map to chromosome Y is greater than 1%, greater than 2%, greater than 3%, greater than 4%, greater than 5%, greater than 6%, greater than 7%, greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, greater than 20%, greater than 25%, or greater than 30%.

In some embodiments, a composition is described com- 10 prising a mixture of DNA of fetal origin, and DNA of maternal origin, wherein the percent of sequences that uniquely map to a chromosome, and that contains at least one single nucleotide polymorphism is greater than 0.2%, greater than 0.3%, greater than 0.4%, greater than 0.5%, 15 greater than 0.6%, greater than 0.7%, greater than 0.8%, greater than 0.9%, greater than 1%, greater than 1.2%, greater than 1.4%, greater than 1.6%, greater than 1.8%, greater than 2%, greater than 2.5%, greater than 3%, greater than 4%, greater than 5%, greater than 6%, greater than 7%, 20 greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, or greater than 20%, and where the chromosome is taken from the group 13, 18, 21, X, or Y. In some embodiments of the present disclosure, there is a and DNA of maternal origin, wherein the percent of sequences that uniquely map to a chromosome and that contain at least one single nucleotide polymorphism from a set of single nucleotide polymorphisms is greater than 0.15%, greater than 0.2%, greater than 0.3%, greater than 300.4%, greater than 0.5%, greater than 0.6%, greater than 0.7%, greater than 0.8%, greater than 0.9%, greater than 1%, greater than 1.2%, greater than 1.4%, greater than 1.6%, greater than 1.8%, greater than 2%, greater than 2.5%, greater than 3%, greater than 4%, greater than 5%, greater 3: than 6%, greater than 7%, greater than 8%, greater than 9%, greater than 10%, greater than 12%, greater than 15%, or greater than 20%, where the chromosome is taken from the set of chromosome 13, 18, 21, X and Y, and where the number of single nucleotide polymorphisms in the set of 40 single nucleotide polymorphisms is between 1 and 10, between 10 and 20, between 20 and 50, between 50 and 100, between 100 and 200, between 200 and 500, between 500 and 1,000, between 1,000 and 2,000, between 2,000 and 5,000, between 5,000 and 10,000, between 10,000 and 45 20,000, between 20,000 and 50,000, and between 50,000 and 100,000.

In theory, each cycle in the amplification doubles the amount of DNA present; however, in reality, the degree of amplification is slightly lower than two. In theory, amplifi- 50 cation, including targeted amplification, will result in bias free amplification of a DNA mixture; in reality, however, different alleles tend to be amplified to a different extent than other alleles

When DNA is amplified, the degree of allelic bias typi- 55 cally increases with the number of amplification steps. In some embodiments, the methods described herein involve amplifying DNA with a low level of allelic bias. Since the allelic bias compounds with each additional cycle, one can determine the per cycle allelic bias by calculating the nth 60 root of the overall bias where n is the base 2 logarithm of degree of enrichment. In some embodiments, there is a composition comprising a second mixture of DNA, where the second mixture of DNA has been preferentially enriched at a plurality of polymorphic loci from a first mixture of 65 DNA where the degree of enrichment is at least 10, at least 100, at least 1,000, at least 10,000, at least 100,000 or at least

202

1,000,000, and where the ratio of the alleles in the second mixture of DNA at each locus differs from the ratio of the alleles at that locus in the first mixture of DNA by a factor that is, on average, less than 1,000%, 500%, 200%, 100%, 50%, 20%, 10%, 5%, 2%, 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01%. In some embodiments, there is a composition comprising a second mixture of DNA, where the second mixture of DNA has been preferentially enriched at a plurality of polymorphic loci from a first mixture of DNA where the per cycle allelic bias for the plurality of polymorphic loci is, on average, less than 10%, 5%, 2%, 1%, 0.5%, 0.2%, 0.1%, 0.05%, or 0.02%. In some embodiments, the plurality of polymorphic loci comprises at least 10 loci, at least 20 loci, at least 50 loci, at least 100 loci, at least 200 loci, at least 500 loci, at least 1,000 loci, at least 2,000 loci, at least 5,000 loci, at least 10,000 loci, at least 20,000 loci, or at least 50,000 loci.

### Some Embodiments

In some embodiments, a method is disclosed herein for generating a report disclosing the determined ploidy status of a chromosome in a gestating fetus, the method comprising: obtaining a first sample that contains DNA from the composition comprising a mixture of DNA of fetal origin, 25 mother of the fetus and DNA from the fetus; obtaining genotypic data from one or both parents of the fetus; preparing the first sample by isolating the DNA so as to obtain a prepared sample; measuring the DNA in the prepared sample at a plurality of polymorphic loci; calculating, on a computer, allele counts or allele count probabilities at the plurality of polymorphic loci from the DNA measurements made on the prepared sample; creating, on a computer, a plurality of ploidy hypotheses concerning expected allele count probabilities at the plurality of polymorphic loci on the chromosome for different possible ploidy states of the chromosome; building, on a computer, a joint distribution model for allele count probability of each polymorphic locus on the chromosome for each ploidy hypothesis using genotypic data from the one or both parents of the fetus; determining, on a computer, a relative probability of each of the ploidy hypotheses using the joint distribution model and the allele count probabilities calculated for the prepared sample; calling the ploidy state of the fetus by selecting the ploidy state corresponding to the hypothesis with the greatest probability; and generating a report disclosing the determined ploidy status.

In some embodiments, the method is used to determine the ploidy state of a plurality of gestating fetuses in a plurality of respective mothers, the method further comprising: determining the percent of DNA that is of fetal origin in each of the prepared samples; and wherein the step of measuring the DNA in the prepared sample is done by sequencing a number of DNA molecules in each of the prepared samples, where more molecules of DNA are sequenced from those prepared samples that have a smaller fraction of fetal DNA than those prepared samples that have a larger fraction of fetal DNA.

In some embodiments, the method is used to determine the ploidy state of a plurality of gestating fetuses in a plurality of respective mothers, and where the measuring the DNA in the prepared sample is done, for each of the fetuses, by sequencing a first fraction of the prepared sample of DNA to give a first set of measurements, the method further comprising: making a first relative probability determination for each of the ploidy hypotheses for each of the fetuses, given the first set of DNA measurements; resequencing a second fraction of the prepared sample from those fetuses

# US 11,519,035 B2

203

where the first relative probability determination for each of the ploidy hypotheses indicates that a ploidy hypothesis corresponding to an aneuploid fetus has a significant but not conclusive probability, to give a second set of measurements; making a second relative probability determination for ploidy hypotheses for the fetuses using the second set of measurements and optionally also the first set of measurements; and calling the ploidy states of the fetuses whose second sample was resequenced by selecting the ploidy state corresponding to the hypothesis with the greatest probability as determined by the second relative probability determination

In some embodiments, a composition of matter is disclosed, the composition of matter comprising: a sample of preferentially enriched DNA, wherein the sample of preferentially enriched DNA has been preferentially enriched at a plurality of polymorphic loci from a first sample of DNA, wherein the first sample of DNA consisted of a mixture of maternal DNA and fetal DNA derived from maternal plasma, where the degree of enrichment is at least a factor of 2, and wherein the allelic bias between the first sample and the preferentially enriched sample is, on average, selected from the group consisting of less than 2%, less than 1%, less than 0.5%, less than 0.2%, less than 0.1%, less than 25 0.05%, less than 0.02%, and less than 0.01%. In some embodiments, a method is disclosed to create a sample of such preferentially enriched DNA.

In some embodiment, a method is disclosed for determining the presence or absence of a fetal aneuploidy in a maternal tissue sample comprising fetal and maternal genomic DNA, wherein the method comprises: (a) obtaining a mixture of fetal and maternal genomic DNA from said maternal tissue sample; (b) selectively enriching the mixture of fetal and maternal DNA at a plurality of polymorphic alleles; (c) distributing selectively enriched fragments from the mixture of fetal and maternal genomic DNA of step a to provide reaction samples comprising a single genomic DNA molecule or amplification products of a single genomic 40 DNA molecule; (d) conducting massively parallel DNA sequencing of the selectively enriched fragments of genomic DNA in the reaction samples of step c) to determine the sequence of said selectively enriched fragments; (e) identifying the chromosomes to which the sequences obtained in 45 step d) belong; (f) analyzing the data of step d) to determine i) the number of fragments of genomic DNA from step d) that belong to at least one first target chromosome that is presumed to be diploid in both the mother and the fetus, and ii) the number of fragments of genomic DNA from step d) 50 that belong to a second target chromosome, wherein said second chromosome is suspected to be aneuploid in the fetus; (g) calculating an expected distribution of the number of fragments of genomic DNA from step d) for the second target chromosome if the second target chromosome is 55 euploid, using the number determined in step f) part i); (h) calculating an expected distribution of the number of fragments of genomic DNA from step d) for the second target chromosome if the second target chromosome is an uploid, using the first number is step f) part i) and an estimated 60 fraction of fetal DNA found in the mixture of step b); and (i) using a maximum likelihood or maximum a posteriori approach to determine whether the number of fragments of genomic DNA determined in step f) part ii) is more likely to be part of the distribution calculated in step g) or the 65 distribution calculated in step h); thereby indicating the presence or absence of a fetal aneuploidy.

204

Exemplary Cancer Diagnostic Methods

Note that it has been demonstrated that DNA that originated from cancer that is living in a host can be found in the blood of the host. In the same way that genetic diagnoses can be made from the measurement of mixed DNA found in maternal blood, genetic diagnoses can equally well be made from the measurement of mixed DNA found in host blood. The genetic diagnoses may include aneuploidy states, or gene mutations. Any claim in the instant disclosure that reads on determining the ploidy state or genetic state of a fetus from the measurements made on maternal blood can equally well read on determining the ploidy state or genetic state of a cancer from the measurements on host blood.

In some embodiments, a method of the present disclosure allows one to determine the ploidy status of a cancer, the method including obtaining a mixed sample that contains genetic material from the host, and genetic material from the cancer; measuring the DNA in the mixed sample; calculating the fraction of DNA that is of cancer origin in the mixed sample; and determining the ploidy status of the cancer using the measurements made on the mixed sample and the calculated fraction. In some embodiments, the method may further include administering a cancer therapeutic based on the determination of the ploidy state of the cancer. In some embodiments, the method may further include administering a cancer therapeutic based on the determination of the ploidy state of the cancer, wherein the cancer therapeutic is taken from the group comprising a pharmaceutical, a biologic therapeutic, and antibody based therapy and combination thereof.

Exemplary Clinical Actions

In some embodiments, any of the methods include taking a clinical action based on a result of a method of the invention (such as the determination of the presence or absence of a polymorphism or mutation, ploidy state, or paternity). In some embodiments in which an embryo or fetus has one or more one or more polymorphisms or mutations of interest (such as a CNV) based on a result of a method of the invention, the clinical action includes performing additional testing (such as testing to confirm the presence of the polymorphism or mutation), not implanting the embryo for IVF, implanting a different embryo for IVF terminating a pregnancy, preparing for a special needs child, or undergoing an intervention designed to decrease the severity of the phenotypic presentation of a genetic disorder. In some embodiments, the clinical action is selected from the group consisting of performing an ultrasound, amniocentesis on the fetus, amniocentesis on a subsequent fetus that inherits genetic material from the mother and/or father, chorion villus biopsy on the fetus, chorion villus biopsy on a subsequent fetus that inherits genetic material from the mother and/or father, in vitro fertilization, preimplantation genetic diagnosis on one or more embryos that inherited genetic material from the mother and/or father, karyotyping on the mother, karyotyping on the father, fetal echocardiogram (such as an echocardiogram of a fetus with trisomy 21, 18, or 13, monosomy X, or a microdeletion), and combinations thereof. In some embodiments, the clinical action is selected from the group consisting of administering growth hormone to a born child with monosomy X (such as administration starting at ~9 months), administering calcium to a born child with a 22q deletion (such as DiGeorge syndrome), administering an androgen such as testosterone to a born child with 47,XXY (such as one injection per month for 3 months of 25 mg testosterone enanthate to an infant or toddler), performing a test for cancer on a woman with a complete or partial molar pregnancy (such as a triploid fetus), administering a therapy for cancer such as a chemo-

# US 11,519,035 B2

205

therapeutic agent to a woman with a complete or partial molar pregnancy (such as a triploid fetus), screening a fetus determined to be male (such as a fetus determined to be male using a method of the invention) for one or more X-linked genetic disorders such as Duchenne muscular dystrophy (DMD), adrenoleukodystrophy, or hemophilia, performing amniocentesis on a male fetus at risk for an X-linked disorder, administering dexamethasone to a women with a female fetus at risk male (such as a fetus determined to be female using a method of the invention) for congenital 10 adrenal hyperplasia, performing amniocentesis on a female fetus at risk for congenital adrenal hyperplasia, administering killed vaccines (instead of live vaccines) or not administering certain vaccines to a born child who is (or is suspected of being) immune deficient from a 22q11.2 dele- 15 tion, performing occupational and/or physical therapy, performing early intervention in education, delivering the baby at a tertiary care center with a NICU and/or having pediatric specialists available at delivery, behavioral intervention for born child (such as a child with XXX, XXY, or XYY), and 20

In some embodiments, ultrasound or another screening test is performed on a women determined to have multiple pregnancies (such as twins) to determine whether or not two or more of the fetus are monochorionic. Monozygotic twins 25 result from ovulation and fertilization of a single oocyte, with subsequent division of the zygote; placentation may be dichorionic or monochorionic. Dizygotic twins occur from ovulation and fertilization of two oocytes, which usually results in dichorionic placentation. Monochorionic twins 30 sample. have a risk of twin-to-twin transfusion syndrome, which may cause unequal distribution of blood between fetuses that results in differences in their growth and development, sometimes resulting in stillbirth. Thus, twins determined to be monozygotic twins using a method of the invention are 35 desirably tested (such as by ultrasound) to determine if they are monochorionic twins, and if so, these twins can be monitored (such as bi-weekly ultrasounds from 16 weeks) for signs of win-to-twin transfusion syndrome.

In some embodiments in which an embryo or fetus does 40 not have one or more one or more polymorphisms or mutations of interest (such as a CNV) based on a result of a method of the invention, the clinical action includes implanting the embryo for IVF or continuing a pregnancy. In some embodiments, the clinical action is additional testing 45 to confirm the absence of the polymorphism or mutation selected from the group consisting of performing an ultrasound, amniocentesis, chorion villus biopsy, and combinations thereof.

In some embodiments in which an individual has one or 50 more polymorphisms or mutations (such as a polymorphism or mutation associated with a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer) based on a result of a method of the invention, the clinical action includes performing additional testing or 55 administering one or more therapies for a disease or disorder (such as a therapy for cancer, a therapy for the specific type of cancer or type of mutation the individual is diagnosed with, or any of the therapies disclosed herein). In some embodiments, the clinical action is additional testing to 60 confirm the presence or absence of a polymorphism or mutation selected from the group consisting of biopsy, surgery, medical imaging (such as a mammogram or an ultrasound), and combinations thereof.

In some embodiments, the additional testing includes 65 performing the same or a different method (such as any of the methods described herein) to confirm the presence or

206

absence of the polymorphism or mutation (such as a CNV), such as testing either a second fraction of the same sample that was tested or a different sample from the same individual (such as the same pregnant mother, fetus, embryo, or individual at increased risk for cancer). In some embodiments, the additional testing is performed for an individual for whom the probability of a polymorphism or mutation (such as a CNV) is above a threshold value. In some embodiments, the additional testing is performed for an individual for whom the confidence or z-score for the determination of a polymorphism or mutation (such as a CNV) is above a threshold value (such as additional testing to confirm the presence of a likely polymorphism or mutation). In some embodiments, the additional testing is performed for an individual for whom the confidence or z-score for the determination of a polymorphism or mutation (such as a CNV) is between minimum and maximum threshold values (such as additional testing to increase the confidence that the initial result is correct). In some embodiments, the additional testing is performed for an individual for whom the confidence for the determination of the presence or absence of a polymorphism or mutation (such as a CNV) is below a threshold value (such as a "no call" result due to not being able to determine the presence or absence of the CNV with sufficient confidence). An exemplary Z core is calculated in Chiu et al. BMJ 2011; 342:c7401 (which is hereby incorporated by reference in its entirety) in which chromosome 21 is used as an example and can be replaced with any other chromosome or chromosome segment in the test

Z score for percentage chromosome 21 in test case=
((percentage chromosome 21 in test case)(mean percentage chromosome 21 in reference
controls))/(standard deviation of percentage
chromosome 21 in reference controls).

In some embodiments, the additional testing is performed for an individual for whom the initial sample did not meet quality control guidelines or had a fetal fraction or a tumor fraction below a threshold value. In some embodiments, the method includes selecting an individual for additional testing based on the result of a method of the invention, the probability of the result, the confidence of the result, or the z-score; and performing the additional testing on the individual (such as on the same or a different sample). In some embodiments, a subject diagnosed with a disease or disorder (such as cancer) undergoes repeat testing using a method of the invention or known testing for the disease or disorder at multiple time points to monitor the progression of the disease or disorder or the remission or reoccurrence of the disease or disorder.

**Exemplary Implementation Methods** 

Any of the embodiments disclosed herein may be implemented in digital electronic circuitry, integrated circuitry, specially designed ASICs (application-specific integrated circuits), computer hardware, firmware, software, or in combinations thereof. Apparatus of the presently disclosed embodiments can be implemented in a computer program product tangibly embodied in a machine-readable storage device for execution by a programmable processor; and method steps of the presently disclosed embodiments can be performed by a programmable processor executing a program of instructions to perform functions of the presently disclosed embodiments by operating on input data and generating output. The presently disclosed embodiments can be implemented advantageously in one or more computer programs that are executable and/or interpretable on a programmable system including at least one programmable

# US 11,519,035 B2

207

processor, which may be special or general purpose, coupled to receive data and instructions from, and to transmit data and instructions to, a storage system, at least one input device, and at least one output device. Each computer program can be implemented in a high-level procedural or object-oriented programming language or in assembly or machine language if desired; and in any case, the language can be a compiled or interpreted language. A computer program may be deployed in any form, including as a stand-alone program, or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program may be deployed to be executed or interpreted on one computer or on multiple computers at one site, or distributed across multiple sites and interconnected by a communication network.

Computer readable storage media, as used herein, refers to physical or tangible storage (as opposed to signals) and includes without limitation volatile and non-volatile, removable and non-removable media implemented in any method or technology for the tangible storage of information such as computer-readable instructions, data structures, program modules or other data. Computer readable storage media includes, but is not limited to, RAM, ROM, EPROM, EEPROM, flash memory or other solid state memory technology, CD-ROM, DVD, or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other physical or material medium which can be used to tangibly store the desired information or data or instructions and which can be accessed by a computer or processor.

In some embodiments, the invention features a computer configured to accomplish one or more of the in vitro methods described herein. In some embodiments, the data is analyzed by the computer system as described herein. In some embodiments, genetic data (such as sequencing or microarray data) from at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different loci is analyzed by the computer is less than 200, 100, 60, 30, 20, 10, 5, or 1 minute, or in less than 30 or 10 seconds to detect the present or absence of a mutation (such as a CNV or SNV) at the loci.

Any of the methods described herein may include the 45 output of data in a physical format, such as on a computer screen, or on a paper printout. In explanations of any embodiments elsewhere in this document, it should be understood that the described methods may be combined with the output of the actionable data in a format that can be 50 acted upon by a physician. In addition, the described methods may be combined with the actual execution of a clinical decision that results in a clinical treatment, or the execution of a clinical decision to make no action. Some of the embodiments described in the document for determining 55 genetic data pertaining to a target individual may be combined with the decision to select one or more embryos for transfer in the context of IVF, optionally combined with the process of transferring the embryo to the womb of the prospective mother. Some of the embodiments described in 60 the document for determining genetic data pertaining to a target individual may be combined with the notification of a potential chromosomal abnormality, or lack thereof, with a medical professional, optionally combined with the decision to abort, or to not abort, a fetus in the context of prenatal 65 diagnosis. Some of the embodiments described herein may be combined with the output of the actionable data, and the

208

execution of a clinical decision that results in a clinical treatment, or the execution of a clinical decision to make no action

Exemplary Diagnostic Boxes

In an embodiment, the present disclosure comprises a diagnostic box that is capable of partly or completely carrying out any of the methods described in this disclosure. In an embodiment, the diagnostic box may be located at a physician's office, a hospital laboratory, or any suitable location reasonably proximal to the point of patient care. The box may be able to run the entire method in a wholly automated fashion, or the box may require one or a number of steps to be completed manually by a technician. In an embodiment, the box may be able to analyze at least the genotypic data measured on the maternal plasma. In an embodiment, the box may be linked to means to transmit the genotypic data measured on the diagnostic box to an external computation facility which may then analyze the genotypic data, and possibly also generate a report. The diagnostic box may include a robotic unit that is capable of transferring aqueous or liquid samples from one container to another. It may comprise a number of reagents, both solid and liquid. It may comprise a high throughput sequencer. It may comprise a computer.

Experimental Section

The presently disclosed embodiments are described in the following Examples, which are set forth to aid in the understanding of the disclosure, and should not be construed to limit in any way the scope of the disclosure as defined in the claims which follow thereafter. The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to use the described embodiments, and are not intended to limit the scope of the disclosure nor are they intended to represent that the Examples below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by volume, and temperature is in degrees Centigrade. It should be understood that variations in the methods as described may be made without changing the fundamental aspects that the Examples are meant to illustrate.

# Example 1

The objective was to show that a Bayesian maximum likelihood estimation (MLE) algorithm that uses parent genotypes to calculate fetal fraction improves accuracy of non-invasive prenatal trisomy diagnosis compared to published methods.

Simulated sequencing data for maternal cfDNA was created by sampling reads obtained on trisomy-21 and respective mother cell lines. The rate of correct disomy and trisomy calls were determined from 500 simulations at various fetal fractions for a published method (Chiu et al. BMJ 2011; 342:c7401) and our MLE-based algorithm. We validated the simulations by obtaining 5 million shotgun reads from four pregnant mothers and respective fathers collected under an IRB-approved protocol. Parental genotypes were obtained on a 290K SNP array. (See FIG. 14)

In simulations, the MLE-based approach achieved 99.0% accuracy for fetal fractions as low as 9% and reported confidences that corresponded well to overall accuracy. We validated these results using four real samples wherein we obtained all correct calls with a computed confidence exceeding 99%. In contrast, our implementation of the

# US 11,519,035 B2

### 209

published algorithm for Chiu et al. required 18% fetal fraction to achieve 99.0% accuracy, and achieved only 87.8% accuracy at 9% fetal DNA.

Fetal fraction determination from parental genotypes in conjunction with a MLE-based approach achieves greater accuracy than published algorithms at the fetal fractions expected during the 1st and early 2nd trimester. Furthermore, the method disclosed herein produces a confidence metric that is crucial in determining the reliability of the result, especially at low fetal fractions where ploidy detection is more difficult. Published methods use a less accurate threshold method for calling ploidy based on large sets of disomy training data, an approach that predefines a false positive rate. In addition, without a confidence metric, published methods are at risk of reporting false negative results when there is insufficient fetal cfDNA to make a call.

In some embodiments, a confidence estimate is calculated for the called ploidy state.

### Example 2

The objective was to improve non-invasive detection of fetal trisomy 18, 21, and X particularly in samples consisting of low fetal fraction by using a targeted sequencing approach combined with parent genotypes and Hapmap data in a Bayesian Maximum Likelihood Estimation (MLE) algo- 25 rithm.

Maternal samples from four euploid and two trisomy-positive pregnancies and respective paternal samples were obtained under an IRB-approved protocol from patients where fetal karyotype was known. Maternal cfDNA was 30 extracted from plasma and roughly 10 million sequence reads were obtained following preferential enrichment that targeted specific SNPs. Parent samples were similarly sequenced to obtain genotypes.

The described algorithm correctly called chromosome 18 and 21 disomy for all euploid samples and normal chromosomes of aneuploid samples. Trisomy 18 and 21 calls were correct, as were chromosome X copy numbers in male and female fetuses. The confidence produced by the algorithm was in excess of 98% in all cases.

The method described accurately reported the ploidy of all tested chromosomes from six samples, including samples comprised of less than 12% fetal DNA, which account for roughly 30% of  $1^{st}$  and early  $2^{nd}$ -trimester samples. The crucial difference between the instant MLE algorithm and 45 published methods is that it leverages parent genotypes and Hapmap data to improve accuracy and generate a confidence metric. At low fetal fractions, all methods become less accurate; it is important to correctly identify samples without sufficient fetal cfDNA to make a reliable call. Others 50 have used chromosome Y specific probes to estimate fetal fraction of male fetuses, but concurrent parental genotyping enables estimation of fetal fraction for both sexes. Another inherent limitation of published methods using untargeted shotgun sequencing is that accuracy of ploidy calling varies 55 among chromosomes due to differences in factors such as GC richness. The instant targeted sequencing approach is largely independent of such chromosome-scale variations and yields more consistent performance between chromosomes.

# Example 3

The objective was to determine if trisomy is detectable with high confidence on a triploid fetus, using novel informatics to analyze SNP loci of free floating fetal DNA in maternal plasma.

### 210

20 mL of blood was drawn from a pregnant patient following abnormal ultrasound. After centrifugation, maternal DNA was extracted from the buffy coat (DNEASY, QIAGEN); cell-free DNA was extracted from plasma (QIAAMP QIAGEN). Targeted sequencing was applied to SNP loci on chromosomes 2, 21, and X in both DNA samples. Maximum-Likelihood Bayesian estimation selected the most likely hypothesis from the set of all possible ploidy states. The method determines fetal DNA fraction, ploidy state and explicit confidences in the ploidy determination. No assumptions are made about the ploidy of a reference chromosome. The diagnostic uses a test statistic that is independent of sequence read counts, which is the recent state of the art.

The instant method accurately diagnosed trisomy of chromosomes 2 and 21. Child fraction was estimated at 11.9% [CI 11.7-12.1]. The fetus was found to have one maternal and two paternal copies of chromosomes 2 and 21 with confidence of effectively 1 (error probability<10<sup>-30</sup>). This was achieved with 92,600 and 258,100 reads on chromosomes 2 and 21 respectively.

This is the first demonstration of non-invasive prenatal diagnosis of trisomic chromosomes from maternal blood where the fetus was triploid, as confirmed by metaphase karyotype. Extant methods of non-invasive diagnosis would not detect aneuploidy in this sample. Current methods rely on a surplus of sequence reads on a trisomic chromosome relative to disomic reference chromosomes; but a triploid fetus has no disomic reference. Furthermore, extant methods would not achieve similarly high-confidence ploidy determination with this fraction of fetal DNA and number of sequence reads. It is straightforward to extend the approach to all 24 chromosomes.

# Example 4

The following protocol was used for 800-plex amplification of DNA isolated from maternal plasma from a euploid pregnancy and also genomic DNA from a triploidy 21 cell 40 line using standard PCR (meaning no nesting was used). Library preparation and amplification involved single tube blunt ending followed by A-tailing. Adaptor ligation was run using the ligation kit found in the AGILENT SURESELECT kit, and PCR was run for 7 cycles. Then, 15 cycles of STA (95° C. for 30 s; 72° C. for 1 min; 60° C. for 4 min; 65° C. for 1 min; 72° C. for 30 s) using 800 different primer pairs targeting SNPs on chromosomes 2, 21 and X. The reaction was run with 12.5 nM primer concentration. The DNA was then sequenced with an ILLUMINA IIGAX sequencer. The sequencer output 1.9 million reads, of which 92% mapped to the genome; of those reads that mapped to the genome, more than 99% mapped to one of the regions targeted by the targeted primers. The numbers were essentially the same for both the plasma DNA and the genomic DNA. FIG. 15 shows the ratio of the two alleles for the ~780 SNPs that were detected by the sequencer in the genomic DNA that was taken from a cell line with known trisomy at chromosome 21. Note that the allele ratios are plotted here for ease of visualization, because the allele distributions are not straightforward to read visually. The circles represent SNPs on disomic chromosomes, while the stars represent SNPs on a trisomic chromosome. FIG. 16 is another representation of the same data as in FIG. 15, where the Y-axis is the relative number of A and B measured for each SNP, and where the X-axis is the SNP number where the SNPs are separated by chromosome. In FIG. 16, SNP 1 to 312 are found on chromosome 2, from SNP 313 to 605 are found on chro-

# US 11,519,035 B2

### 21

mosome 21 which is trisomic, and from SNP 606 to 800 are on chromosome X. The data from chromosomes 2 and X show a disomic chromosome, as the relative sequence counts lie in three clusters: AA at the top of the graph, BB at the bottom of the graph, and AB in the middle of the graph. The data from chromosome 21, which is trisomic, shows four clusters: AAA at the top of the graph, AAB around the 0.65 line (2/3), ABB around the 0.35 line (1/3), and BBB at the bottom of the graph.

FIGS. 17A-D show data for the same 800-plex protocol, but measured on DNA that was amplified from four plasma samples from pregnant women. For these four samples, we expect to see seven clusters of dots: (1) along the top of the graph are those loci where both the mother and the fetus are AA, (2) slightly below the top of the graph are those loci where the mother is AA and the fetus is AB, (3) slightly above the 0.5 line are those loci where the mother is AB and the fetus is AA, (4) along the 0.5 line are those loci where the mother and the fetus are both AB, (5) slightly below the 20 0.5 line are those loci where the mother is AB and the fetus is BB. (6) slightly above the bottom of the graph are those loci where the mother is BB and the fetus is AB, (1) along the bottom of the graph are those loci where both the mother and the fetus are BB. The smaller the fetal fraction, the less  $\,^{25}$ the separation between clusters (1) and (2), between clusters (3), (4) and (5), and between clusters (6) and (7). The separation is expected to be half of the fraction of DNA that is of fetal origin. For example if the DNA is 20% fetal, and 80% maternal, we expect (1) through (7) to be centered at 1.0, 0.9, 0.6, 0.5, 0.4, 0.1 and 0.0 respectively; see for example FIG. 17D, POOL1\_BC5\_ref\_rate. If, instead the DNA is 8% fetal, and 92% maternal, we expect (1) through (7) to be centered at 1.00, 0.96, 0.54, 0.50, 0.46, 0.04 and 0.00 respectively; see for example FIG. 17B, POOL1\_BC2\_ref\_rate. If there is not fetal DNA detected, we do not expect to see (2), (3), (5), or (6); alternately we could say that the separation is zero, and therefore (1) and (2) are on top of each other, as are (3), (4) and (5), and also  $_{40}$ (6) and (7); see e.g. FIG. 17C, POOL1\_BC7\_ref\_rate. Note that the fetal fraction for FIG. 17A, POOL1\_BC1\_ref\_rate is about 25%.

# Example 5

Most methods of DNA amplification and measurement will produce some allele bias, wherein the two alleles that are typically found at a locus are detected with intensities or counts that are not representative of the actual amounts of 50 alleles in the sample of DNA. For example, for a single individual, at a heterozygous locus we expect to see a 1:1 ratio of the two alleles, which is the theoretical ratio expected for a heterozygous locus; however due to allele bias, we may see 55:45, or even 60:40. Also note that in the 55 context of sequencing, if the depth of read is low, then simple stochastic noise could result in significant allele bias. In an embodiment, it is possible to model the behavior of each SNP such that if a consistent bias is observed for particular alleles, this bias can be corrected for. FIG. 18 60 shows the fraction of data that can be explained by binomial variance, before and after bias correction. In FIG. 18, the stars represent the observed allele bias on raw sequence data for the 800-plex experiment; the circles represent the allele bias after correction. Note that if there were no allele bias at 65 all, we would expect the data to fall along the x=y line. A similar set of data that was produced by amplifying DNA

### 212

using a 150-plex targeted amplification produced data that fell very closely on the 1:1 line after bias correction.

### Example 6

Universal amplification of DNA using ligated adaptors with primers specific to the adaptor tags, where the primer annealing and extension times are limited to a few minutes has the effect of enriching the proportion of shorter DNA strands. Most library protocols designed for creating DNA libraries suitable for sequencing contain such a step, and example protocols are published and well known to those in the art. In some embodiments of the invention, adaptors with a universal tag are ligated to the plasma DNA, and amplified using primers specific to the adaptor tag. In some embodiments, the universal tag can be the same tag as used for sequencing, it can be a universal tag only for PCR amplification, or it can be a set of tags. Since the fetal DNA is typically short in nature, while the maternal DNA can be both short and long in nature, this method has the effect of enriching the proportion of fetal DNA in the mixture. The free floating DNA, thought to be DNA from apoptotic cells, and which contains both fetal and maternal DNA, is shortmostly under 200 bp. Cellular DNA released by cell lysis, a common phenomenon after phlebotomy, is typically almost exclusively maternal, and is also quite long-mostly above 500 bp. Therefore, blood samples that have sat around for more than a few minutes will contain a mixture of short (fetal+maternal) and longer (maternal) DNA. Performing a universal amplification with relatively short extension times on maternal plasma followed by targeted amplification will tend to increase the relative proportion of fetal DNA when compared to the plasma that has been amplified using targeted amplification alone. This can be seen in FIG. 19 which shows the measured fetal percent when the input is plasma DNA (vertical axis) vs. the measured fetal percent when the input DNA is plasma DNA that has had a library prepared using the ILLUMINA GAIIx library preparation protocol. All the dots fall below the line, indicating that the library preparation step enriches the fraction of DNA that is of fetal origin. Two samples of plasma that were red, indicating hemolysis and therefore that there would be an increased amount of long maternal DNA present from cell lysis, show a particularly significant enrichment of fetal 45 fraction when the library preparation is performed prior to targeted amplification. The method disclosed herein is particularly useful in cases where there is hemolysis or some other situation has occurred where cells comprising relatively long strands of contaminating DNA have lysed, contaminating the mixed sample of short DNA with the long DNA. Typically the relatively short annealing and extension times are between 30 seconds and 2 minutes, though they could be as short as 5 or 10 seconds or less, or as long as 5 or 10 minutes.

# Example 7

The following protocol was used for 1,200-plex amplification of DNA isolated from maternal plasma from a euploid pregnancy and also genomic DNA from a triploidy 21 cell line using a direct PCR protocol, and also a semi-nested approach. Library preparation and amplification involved single tube blunt ending followed by A-tailing. Adaptor ligation was run using a modification of the ligation kit found in the AGILENT SURESELECT kit, and PCR was run for 7 cycles. In the targeted primer pool, there were 550 assays for SNPs from chromosome 21, and 325 assays for

Case: 24-1324 Document: 42-1 Page: 229 Filed: 03/18/2024

# US 11,519,035 B2

### 213

SNPs from each of chromosomes 1 and X. Both protocols involved 15 cycles of STA (95° C. for 30 s; 72° C. for 1 min; 60° C. for 4 min; 65° C. for 30 s; 72° C. for 30 s) using 16 nM primer concentration. The semi-nested PCR protocol involved a second amplification of 15 cycles of STA (95° C. for 30 s; 72° C. for 1 min; 60° C. for 4 min; 65° C. for 30 s; 72° C. for 30 s) using an inner forward tag concentration of 29 nM, and a reverse tag concentration of 1 uM or 0.1 uM. The DNA was then sequenced with an ILLUMINA IIGAX sequencer. For the direct PCR protocol, 73% of the reads map to the genome; for the semi-nested protocol, 97.2% of the sequence reads map to the genome. Therefore, the semi-nested protocol result in approximately 30% more information, presumably mostly due to the elimination of  $_{15}$ primers that are most likely to cause primer dimers.

The depth of read variability tends to be higher when using the semi-nested protocol than when the direct PCR protocol is used (see FIG. 20) where the diamonds refer to and the squares refer to the depth of read for loci run with no nesting. The SNPs are arranged by depth of read for the diamonds, so the diamonds all fall on a curved line, while the squares appear to be loosely correlated; the arrangements of the SNPs is arbitrary, and it is the height of the dot that 25 denotes depth of read rather than its location left to right.

In some embodiments, the methods described herein can achieve excellent depth of read (DOR) variances. For example, in one version of this Example (FIG. 21) using a 1,200-plex direct PCR amplification of genomic DNA, of 30 the 1,200 assays: 1186 assays had a DOR greater than 10; the average depth of read was 400; 1063 assays (88.6%) had a depth of read of between 200 and 800, and ideal window where the number of reads for each allele is high enough to give meaningful data, while the number of reads for each 35 allele is not so high that the marginal use of those reads was particularly small. Only 12 alleles had higher depth of read with the highest at 1035 reads. The standard deviation of the DOR was 290, the average DOR was 453, the coefficient of variance of the DOR was 64%, there were 950,000 total 40 reads, and 63.1% of the reads mapped to the genome. In another experiment (FIG. 22) using a 1,200-plex seminested protocol, the DOR was higher. The standard deviation of the DOR was 583, the average DOR was 630, the coefficient of variance of the DOR was 93%, there were 45 870,000 total reads, and 96.3% of the reads mapped to the genome. Note, in both these cases, the SNPs are arranged by the depth of read for the mother, so the curved line represents the maternal depth of read. The differentiation between child and father is not significant; it is only the trend that is 50 significant for the purpose of this explanation.

## Example 8

In an experiment, the semi-nested 1,200-plex PCR pro- 55 tocol was used to amplify DNA from one cell and from three cells. This experiment is relevant to prenatal aneuploidy testing using fetal cells isolated from maternal blood, or for preimplantation genetic diagnosis using biopsied blastomeres or trophectoderm samples. There were 3 replicates of 60 1 and 3 cells from 2 individuals (46 XY and 47 XX+21) per condition. Assays targeted chromosomes 1, 21 and X. Three different lysis methods were used: ARCTURUS, MPERv2 and Alkaline lysis. Sequencing was run multiplexing 48 samples in one sequencing lane. The algorithm returned 65 correct ploidy calls for each of the three chromosomes, and for each of the replicates.

# 214

### Example 9

In one experiment, four maternal plasma samples were prepared and amplified using a hemi-nested 9,600-plex protocol. The samples were prepared in the following way: Up to 40 mL of maternal blood were centrifuged to isolate the buffy coat and the plasma. The genomic DNA in the maternal sample was prepared from the buffy coat and paternal DNA was prepared from a blood sample or saliva sample. Cell-free DNA in the maternal plasma was isolated using the QIAGEN CIRCULATING NUCLEIC ACID kit and eluted in 45 uL TE buffer according to manufacturer's instructions. Universal ligation adapters were appended to the end of each molecule of 35 uL of purified plasma DNA and libraries were amplified for 7 cycles using adaptor specific primers. Libraries were purified with AGEN-COURT AMPURE beads and eluted in 50 ul water.

3 ul of the DNA was amplified with 15 cycles of STA (95° C. for 10 min for initial polymerase activation, then 15 the depth of read for loci run with the semi-nested protocol, 20 cycles of 95° C. for 30 s; 72° C. for 10 s; 65° C. for 1 min; 60° C. for 8 min; 65° C. for 3 min and 72° C. for 30 s; and a final extension at 72° C. for 2 min) using 14.5 nM primer concentration of 9600 target-specific tagged reverse primers and one library adaptor specific forward primer at 500 nM.

The hemi-nested PCR protocol involved a second amplification of a dilution of the first STAs product for 15 cycles of STA (95° C. for 10 min for initial polymerase activation, then 15 cycles of 95° C. for 30 s; 65° C. for 1 min; 60° C. for 5 min; 65° C. for 5 min and 72° C. for 30 s; and a final extension at 72° C. for 2 min) using reverse tag concentration of 1000 nM, and a concentration of 16.6 u nM for each of 9600 target-specific forward primers.

An aliquot of the STA products was then amplified by standard PCR for 10 cycles with 1 uM of tag-specific forward and barcoded reverse primers to generate barcoded sequencing libraries. An aliquot of each library was mixed with libraries of different barcodes and purified using a spin column.

In this way, 9,600 primers were used in the single-well reactions; the primers were designed to target SNPs found on chromosomes 1, 2, 13, 18, 21, X and Y. The amplicons were then sequenced using an ILLUMINA GAIIX sequencer. Per sample, approximately 3.9 million reads were generated by the sequencer, with 3.7 million reads mapping to the genome (94%), and of those, 2.9 million reads (74%) mapped to targeted SNPs with an average depth of read of 344 and a median depth of read of 255. The fetal fraction for the four samples was found to be 9.9%, 18.9%, 16.3%, and 21.2%

Relevant maternal and paternal genomic DNA samples amplified using a semi-nested 9600-plex protocol and sequenced. The semi-nested protocol is different in that it applies 9,600 outer forward primers and tagged reverse primers at 7.3 nM in the first STA. Thermocycling conditions and composition of the second STA, and the barcoding PCR were the same as for the hemi-nested protocol.

The sequencing data was analyzed using informatics methods disclosed herein and the ploidy state was called at six chromosomes for the fetuses whose DNA was present in the 4 maternal plasma samples. The ploidy calls for all 28 chromosomes in the set were called correctly with confidences above 99.2% except for one chromosome that was called correctly, but with a confidence of 83%.

FIG. 23 shows the depth of read of the 9,600-plex hemi-nesting approach along with the depth of read of the 1,200-plex semi-nested approach described in Example 7, though the number of SNPs with a depth of read greater than

# US 11,519,035 B2

### 215

100, greater than 200 and greater than 400 was significantly higher than in the 1,200-plex protocol. The number of reads at the 90<sup>th</sup> percentile can be divided by the number of reads at the 10<sup>th</sup> percentile to give a dimensionless metric that is indicative of the uniformity of the depth of read; the smaller the number, the more uniform (narrow) the depth of read. The average 90<sup>th</sup> percentile/10<sup>th</sup> percentile ratio is 11.5 for the method run in Example 9, while it is 5.6 for the method run in Example 7. A narrower depth of read for a given protocol plexity is better for sequencing efficiency, as fewer sequence reads are necessary to ensure that a certain percentage of reads are above a read number threshold.

### Example 10

In one experiment, four maternal plasma samples were prepared and amplified using a semi-nested 9,600-plex protocol. Details of Example 10 were very similar to Example 9, the exception being the nesting protocol, and including the identity of the four samples. The ploidy calls for all 28 chromosomes in the set were called correctly with confidences above 99.7%. 7.6 million (97%) of reads mapped to the genome, and 6.3 million (80%) of the reads mapped to the targeted SNPs. The average depth of read was 751, and the median depth of read was 396.

### Example 11

In one experiment, three maternal plasma samples were split into five equal portions, and each portion was amplified using either 2,400 multiplexed primers (four portions) or 1,200 multiplexed primers (one portion) and amplified using a semi-nested protocol, for a total of 10,800 primers. After amplification, the portions were pooled together for sequencing. Details of Example 11 were very similar to Example 9, the exception being the nesting protocol, and the split and pool approach. The ploidy calls for all 21 chromosomes in the set were called correctly with confidences above 99.7%, except for one missed call where the confidence was 83%. 3.4 million reads mapped to targeted SNPs, the average depth of read was 404 and the median depth of read was 258.

### Example 12

In one experiment, four maternal plasma samples were split into four equal portions, and each portion was amplified using 2,400 multiplexed primers and amplified using a semi-nested protocol, for a total of 9,600 primers. After amplification, the portions were pooled together for sequencing. Details of Example 12 were very similar to Example 9, the exception being the nesting protocol, and the split and pool approach. The ploidy calls for all 28 chromosomes in the set were called correctly with confidences above 97%, except for one missed call where the confidence 55 was 78%. 4.5 million reads mapped to targeted SNPs, the average depth of read was 535 and the median depth of read was 412.

### Example 13

In one experiment, four maternal plasma samples were prepared and amplified using a 9,600-plex triply heminested protocol, for a total of 9,600 primers. Details of Example 12 were very similar to Example 9, the exception 65 being the nesting protocol which involved three rounds of amplification; the three rounds involved 15, 10 and 15 STA

### 216

cycles respectively. The ploidy calls for 27 of 28 chromosomes in the set were called correctly with confidences above 99.9%, except for one that was called correctly with 94.6%, and one missed call with a confidence of 80.8%. 3.5 million reads mapped to targeted SNPs, the average depth of read was 414 and the median depth of read was 249.

### Example 14

In one Example 45 sets of cells were amplified using a 1,200-plex semi-nested protocol, sequenced, and ploidy determinations were made at three chromosomes. Note that this experiment is meant to simulate the conditions of performing pre-implantation genetic diagnosis on single-cell biopsies from day 3 embryos, or trophectoderm biopsies from day 5 embryos. 15 individual single cells and 30 sets of three cells were placed in 45 individual reaction tubes for a total of 45 reactions where each reaction contained cells from only one cell line, but the different reactions contained cells from different cell lines. The cells were prepared into 5 ul washing buffer and lysed the by adding 5 ul ARCTURUS PICOPURE lysis buffer (APPLIED BIOSYSTEMS) and incubating at 56° C. for 20 min, 95° C. for 10 min.

The DNA of the single/three cells was amplified with 25 cycles of STA (95° C. for 10 min for initial polymerase activation, then 25 cycles of 95° C. for 30 s; 72° C. for 10 s; 65° C. for 1 min; 60° C. for 8 min; 65° C. for 3 min and 72° C. for 30 s; and a final extension at 72° C. for 2 min) using 50 nM primer concentration of 1200 target-specific 30 forward and tagged reverse primers.

The semi-nested PCR protocol involved three parallel second amplification of a dilution of the first STAs product for 20 cycles of STA (95° C. for 10 min for initial polymerase activation, then 15 cycles of 95° C. for 30 s; 65° C. for 1 min; 60° C. for 5 min; 65° C. for 5 min and 72° C. for 30 s; and a final extension at 72° C. for 2 min) using reverse tag specific primer concentration of 1000 nM, and a concentration of 60 nM for each of 400 target-specific nested forward primers. In the three parallel 400-plex reactions the total of 1200 targets amplified in the first STA were thus amplified

An aliquot of the STA products was then amplified by standard PCR for 15 cycles with 1 uM of tag-specific forward and barcoded reverse primers to generate barcoded sequencing libraries. An aliquot of each library was mixed with libraries of different barcodes and purified using a spin column.

In this way, 1,200 primers were used in the single cell reactions; the primers were designed to target SNPs found on chromosomes 1, 21 and X. The amplicons were then sequenced using an ILLUMINA GAIIX sequencer. Per sample, approximately 3.9 million reads were generated by the sequencer, with 500,000 to 800,000 million reads mapping to the genome (74% to 94% of all reads per sample).

Relevant maternal and paternal genomic DNA samples from cell lines were analyzed using the same semi-nested 1200-plex assay pool with a similar protocol with fewer cycles and 1200-plex second STA, and sequenced.

The sequencing data was analyzed using informatics methods disclosed herein and the ploidy state was called at the three chromosomes for the samples.

FIG. 24 shows normalized depth of read ratios (vertical axis) for six samples at three chromosomes (1=chrom 1; 2=chrom 21; 3=chrom X). The ratios were set to be equal to the number of reads mapping to that chromosome, normalized, and divided by the number of reads mapping to that chromosome averaged over three wells each comprising

Case: 24-1324 Document: 42-1 Page: 231 Filed: 03/18/2024

# US 11,519,035 B2

### 217

three 46XY cells. The three sets of data points corresponding to the 46XY reactions are expected to have ratios of 1:1. The three sets of data points corresponding to the 47XX+21 cells are expected to have ratios of 1:1 for chromosome 1, 1.5:1 for chromosome 21, and 2:1 for chromosome X.

FIG. 25A-25C show allele ratios plotted for three chromosomes (1, 21, X) for three reaction. The reaction in the lower left shows a reaction on three 46XY cells (FIG. 25B). The left region are the allele ratios for chromosome 1, the middle region are the allele ratios for chromosome 21, and the right region are the allele ratios for chromosome X. For the 46XY cells, for chromosome 1 we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB and BB SNP genotypes. For the 46XY cells, for chromosome 21 we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB and BB SNP genotypes. For the 46XY cells, for chromosome X we expect to see ratios of 1 and 0, corresponding to A, and B SNP genotypes. The reaction in the lower right allele ratios are segregated by chromosome as in the lower left graph. For the 47XX+21 cells, for chromosome 1 we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB and BB SNP genotypes. For the 47XX+21 cells, for chromosome 21 we expect to see ratios of 1, 0.67, 0.33 and 0,  $^{25}$ corresponding to AAA, AAB, ABB and BBB SNP genotypes. For the 47XX+21 cells, for chromosome X we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB, and BB SNP genotypes. The plot in the upper right was made on a reaction comprising 1 ng of genomic DNA from the 30 47XX+21 cell line (FIG. 25A). FIGs. and 26B shows the same graphs as in FIG. 25A-25C, but for reactions performed on only one cell. The left graph was a reaction that contained a 47XX+21 cell (FIG. 26A), and the right graph was for a reaction that contained a 46XX cell (FIG. 26B).

From the graphs shown in FIGS. 25A-25C and FIGS. 26A and 26B, it is visually apparent that there are two clusters of dots for chromosomes where we expect to see ratios of 1 and 0; three clusters of dots for chromosomes where we expect to see ratios of 1, 0.5, and 0, and four clusters of dots for 40 chromosomes where we expect to see ratios of 1, 0.67, 0.33 and 0. The parental support algorithm was able to make correct calls on all of the three chromosomes for all of the 45 reactions.

# Example 15

In one experiment, maternal plasma samples were prepared and amplified using a hemi-nested 19,488-plex protocol. The samples were prepared in the following way: up 50 to 20 mL of maternal blood were centrifuged to isolate the buffy coat and the plasma. The genomic DNA in the maternal sample was prepared from the buffy coat and paternal DNA was prepared from a blood sample or saliva sample. Cell-free DNA in the maternal plasma was isolated using the 55 QIAGEN CIRCULATING NUCLEIC ACID kit and eluted in 50 uL TE buffer according to manufacturer's instructions. Universal ligation adapters were appended to the end of each molecule of 40 uL of purified plasma DNA and libraries were amplified for 9 cycles using adaptor specific primers. 60 Libraries were purified with AGENCOURT AMPURE beads and eluted in 50 ul DNA suspension buffer.

6 ul of the DNA was amplified with 15 cycles of STAR 1 (95° C. for 10 min for initial polymerase activation, then 15 cycles of 96° C. for 30 s; 65° C. for 1 min; 58° C. for 6 65 min;  $60^{\circ}$  C. for 8 min;  $65^{\circ}$  C. for 4 min and  $72^{\circ}$  C. for 30 s; and a final extension at 72° C. for 2 min) using 7.5 nM

### 218

primer concentration of 19,488 target-specific tagged reverse primers and one library adaptor specific forward primer at 500 nM.

The hemi-nested PCR protocol involved a second amplification of a dilution of the STAR 1 product for 15 cycles (STAR 2) (95° C. for 10 min for initial polymerase activation, then 15 cycles of 95° C. for 30 s; 65° C. for 1 min; 60° C. for 5 min;  $65^{\circ}$  C. for 5 min and  $72^{\circ}$  C. for 30 s; and a final extension at 72° C. for 2 min) using reverse tag concentration of 1000 nM, and a concentration of 20 nM for each of 19,488 target-specific forward primers.

An aliquot of the STAR 2 products was then amplified by standard PCR for 12 cycles with 1 uM of tag-specific forward and barcoded reverse primers to generate barcoded sequencing libraries. An aliquot of each library was mixed with libraries of different barcodes and purified using a spin column.

In this way, 19,488 primers were used in the single-well shows a reaction on three 47XX+21 cells (FIG. 25C). The 20 reactions; the primers were designed to target SNPs found on chromosomes 1, 2, 13, 18, 21, X and Y. The amplicons were then sequenced using an ILLUMINA GAIIX sequencer. For plasma samples, approximately 10 million reads were generated by the sequencer, with 9.4-9.6 million reads mapping to the genome (94-96%), and of those, 99.95% mapped to targeted SNPs with a mean depth of read of 460 and a median depth of read of 350. For comparison, a perfectly even distribution would be: 10M reads/19,488 targets=513 reads/target. For primer-dimers, 30,000 reads were from sequenced primer-dimers (0.3% of the reads generated by the sequencer). For genomic samples, 99.4-99.7% of the reads mapped to the genome, of those, 99.99% of the mapped to targeted SNPs, and 0.1% of the reads generated by the sequencer were primer-dimers.

For plasma samples with 10 million sequencing reads, typically at least 19,350 of the 19,488 targeted SNPs (99.3%) are amplified and sequenced. For DNA samples with 2M sequencing reads, typically at least 19,000 targeted SNPs (97.5%) are amplified and sequenced. The lower number may be due to sampling noise since the number of reads is lower and the sequencer misses some of the amplified products. If desired, the number of sequencing reads can be increased to increase the number of targeted SNPs that are amplified and sequenced.

Relevant maternal and paternal genomic DNA samples amplified using a semi-nested 19,488 outer forward primers and tagged reverse primers at 7.5 nM in the STAR 1. Thermocycling conditions and composition of STAR 2, and the barcoding PCR were the same as for the hemi-nested protocol.

The average fetal fraction for 407 samples was found to be 14.8%. The sequencing data was analyzed using informatics methods disclosed herein and the ploidy state was called at four chromosomes (13, 18, 21, Y) for the fetuses whose DNA was present in 378 of the 407 maternal plasma samples, and at chromosome X in 375 of the 407 maternal plasma samples. The ploidy calls for all 1,887 chromosomes in the set were called correctly with confidences above 90%. 1882 of the 1887 calls were above 95%; and 1,862 of the 1,887 calls were called with confidences above 99%.

A similar control experiment was performed using water instead of DNA extracted from plasma in the plasma PCR protocol. Based on six such trials of an experiment, 5-6% of the sequenced reads were primer-dimers. Other sequenced reads were due to background noise. This experiment demonstrates that even in the absence of a nucleic acid sample with target loci for the primers to hybridize to (rather than

# US 11,519,035 B2

### 219

hybridizing to other primers and forming amplified primer dimers) few primer dimers are formed.

### Example 16

The following Example illustrates an exemplary method for designing and selecting a library of primers that can be used in any of the multiplexed PCR methods of the invention. The goal is to select primers from an initial library of candidate primers that can be used to simultaneously 10 amplify a large number of target loci (or a subset of target loci) in a single reaction. For an initial set of candidate target loci, primers did not have to be designed or selected for each target locus. Preferably, primers are designed and selected for a large portion of the most desirable target loci.

15 Step 1

A set of candidate target loci (such as SNPs) were selected based on publically available information about desired parameters for the target loci, such as frequency of the SNPs within a target population or heterozygosity rate of the SNPs 20 (worldwide web at ncbi.nlm.nih.gov/projects/SNP/; Sherry S T, Ward M H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan. 1; 29(1):308-11, which are each incorporated by reference in its entirety). For each candidate locus, one or more PCR 25 primer pairs were designed using the Primer3 program (the worldwide web at primer3.sourceforge.net; libprimer3 release 2.2.3, which is hereby incorporated by reference in its entirety). If there were no feasible designs for PCR primers for a particular target locus, then that target locus 30 was eliminated from further consideration. If desired, a "target locus score" (higher score representing higher desirability) can be calculated for most or all of the target loci, such as a target locus score calculated based on a weighted average of various desired parameters for the target loci. The 35 parameters may be assigned different weights based on their importance for the particular application that the primers will be used for. Exemplary parameters include the heterozygosity rate of the target locus, the disease prevalence associated with a sequence (e.g., a polymorphism) at the target 40 locus, the disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, the specificity of the candidate primer(s) used to amplify the target locus, the size of the candidate primer(s) used to amply the target locus, and the size of the target amplicon. In some embodi- 45 ments, the specificity of the candidate primer for the target locus includes the likelihood that the candidate primer will mis-prime by binding and amplifying a locus other than the target locus it was designed to amplify. In some embodiments, one or more or all the candidate primers that mis- 50 prime are removed from the library. Step 2

À thermodynamic interaction score was calculated between each primer and all primers for all other target loci from Step 1 (see, e.g., Allawi, H. T. & SantaLucia, J., Jr. 55 (1998), "Thermodynamics of Internal C-T Mismatches in DNA", *Nucleic Acids Res.* 26, 2694-2701; Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. (1999), "Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A-A, C-C, G-G, and T-T Mismatches", *Biochemistry* 38, 3468-3477; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest-Neighbor Thermodynamics of Internal A-C Mismatches in DNA: Sequence Dependence and pH Effects", *Biochemistry* 37, 9435-9444.; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA", *Biochemistry* 37, 2170-2179; and Allawi,

### 220

H. T. & SantaLucia, J., Jr. (1997), "Thermodynamics and NMR of Internal G-T Mismatches in DNA", *Biochemistry* 36, 10581-10594; MultiPLX 2.1 (Kaplinski L, Andreson R, Puurand T, Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005 Apr. 15; 21(8):1701-2, which are each hereby incorporated by reference in its entirety). This step resulted in a 2D matrix of interaction scores. The interaction score predicted the likelihood of primer-dimers involving the two interacting primers. The score was calculated as follows:

interaction\_score=max(-delta $G_2$ ,0.8\*(-delta $G_1$ ))

where

deltaG\_2=Gibbs energy (energy required to break the dimer) for a dimer that is extensible by PCR on both ends, i.e., the 3' end of each primer anneals to the other primer; and deltaG\_1=Gibbs energy for a dimer that is extensible by PCR on at least one end.

Step 3:

For each target locus, if there was more than one primerpair design, then one design was selected using the following method:

For each primer-pair design for the locus, find the worstcase (highest) interaction score for the two primers in that design and all primers from all designs for all other target loci.

Pick the design with the best (lowest) worst-case interaction score.

Step 4

A graph was built such that each node represented one locus and its associated primer-pair design (e.g., a Maximal Clique problem). One edge was created between every pair of nodes. A weight was assigned to each edge equal to the worst-case (highest) interaction score between the primers associated with the two nodes connected by the edge. Step 5

If desired, for every pair of designs for two different target loci where one of the primers from one design and one of the primers from the other design would anneal to overlapping target regions, an additional edge was added between the nodes for the two design. The weight of these edges was set equal to the highest weight assigned in Step 4. Thus, Step 5 prevents the library from having primers that would anneal to overlapping target regions, and thus interfere with each other during a multiplex PCR reaction.

An initial interaction score threshold was calculated as follows:

weight\_threshold=max(edge\_weight)-0.05\*(max (edge\_weight)-min(edge\_weight))

where

 $max(edge\_weight)$  is the maximum edge weight in the graph; and

min(edge\_weight) is the minimum edge weight in the graph.

The initial bounds for the threshold were set as follows:

max\_weight\_threshold=max(edge\_weight)

min\_weight\_threshold=min(edge\_weight)

Step 7

Dependence and pH Effects", *Biochemistry* 37, 9435-9444.; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA", *Biochemistry* 37, 2170-2179; and Allawi,

# US 11,519,035 B2

Step 8

Nodes (and all of the edges connected to the removed nodes) were removed from the graph of Step 7 until there were no edges left. Nodes were removed by applying the following procedure repeatedly:

221

Find the node with the highest degree (highest number of edges). If there is more than one then pick one arbitrarily.

Define the set of nodes consisting of the node picked above and all of the nodes connected to it, but excluding any nodes that have degree less than the node picked above.

Choose the node from the set that has the lowest target locus score (lower score representing lower desirability) from Step 1. Remove that node from the graph. Step 9

If the number of nodes remaining in the graph satisfies the 15 required number of target loci for the multiplexed PCR pool (within an acceptable tolerance), then the method was continued at Step 10.

If there were too many or too few nodes remaining in the graph, then a binary search was performed to determine 20 what threshold values would result in the desired number of nodes remaining in the graphs. If there were too many nodes in the graph then, the weight threshold bounds were adjusted as follows:

max\_weight\_threshold=weight\_threshold

Otherwise (if there are two few nodes in the graph), then the weight threshold bounds were adjusted as follows:

min\_weight\_threshold=weight\_threshold

Then, the weight threshold was adjusted follows:

 $weight\_threshold = (max\_weight\_threshold + min\_weight\_threshold)/2$ 

Steps 7-9 were repeated. Step 10

The primer-pair designs associated with the nodes remaining in the graph were selected for the library of primers. This primer library can be used in any of the methods of the invention.

If desired, this method of designing and selecting primers can be performed for primer libraries in which only one primer (instead of a primer pair) is used for amplification of a target locus. In this case, a node presents one primer per target locus (rather than a primer pair).

### Example 17

FIG. 27 is a graph comparing two primer libraries designed using the methods of the invention. This graph 50 shows the number of loci with a particular minor allele frequency that are targeted by each primer library. During the selection of the "new pool" library, more primers were retained. This library enables the amplification of more target loci, especially target loci with relatively large minor 55 allele frequencies (which are the more informative alleles for some method of the invention, such as for detecting fetal chromosomal abnormalities).

These primer libraries were used in the following multiplex PCR method. Blood (20-40 mL) was collected from 60 each subject into two to four CELL-FREE™ DNA tubes (Streck). Plasma (a minimum of 7 mL) was isolated from each sample via a double centrifugation protocol of 2,000 g for 20 min, followed by 3,220 g for 30 min, with supernatant transfer following the first spin. cfDNA was isolated from 65 7-20 mL plasma using the QIAGEN QIAamp Circulating Nucleic Acid kit and eluted in 45 uL TE buffer. Pure

222

maternal genomic DNA was isolated from the buffy coat obtained following the first centrifugation, and pure paternal genomic DNA was prepared similarly from a blood, saliva or buccal sample.

Maternal cfDNA, maternal genomic DNA, and paternal genomic DNA samples were pre-amplified for 15 cycles using 11,000 target-specific assays and an aliquot was transferred to a second PCR reaction of 15 cycles using nested primers. Finally, samples were prepared for sequencing by adding barcoded tags in a third 12-cycle round of PCR. Thus, 11,000 targets were amplified in a single reaction; the targets included SNPs found on chromosomes 13, 18, 21, X, and Y. The amplicons were then sequenced using an ILLU-MINA GAIIx or HISEQ sequencer. Parental genotypes were sequenced at a lower read depth (~20% of cfDNA read depth) than the fetal genotypes.

### Example 18

If desired, the size and quantity of the PCR products can be analyzed using standard methods, such as the use of the Agilent Technologies 2100 Bioanalyzer (FIG. 28A-M). For example, direct PCR methods described herein without 19,488-plex experiments (FIGS. 28B-28G) and 19,488-plex experiments (FIGS. 28B-28D). The amount of primer was 10 nM for FIGS. 28B-28D and 28H to 28J. The amount of primer was 1 nM for FIGS. 28E-28G and 28K to 28M. The amount of input DNA was 24 ng for FIGS. 28B, 28B, 28H, and 28K; 80 ng for FIGS. 28C, 28F, 28, and 28L; and 250 ng for FIGS. 28D, 28G, 28J, and 28M. More input DNA resulted in a greater proportion of the desired 180 base pair product. The peak at 140 base pairs is a primer dimer product.

# Example 19

A proof-of-principle study demonstrated the detection of T13, T18, T21, 45,X, and 47,XXY with equally high accuracies across all chromosomes.

Patients

Pregnant couples were enrolled at specific prenatal care centers under protocols approved by an Institutional Review 45 Board pursuant to local laws. Inclusion criteria were at least 18 years of age, gestational age of at least nine weeks, singleton pregnancies, and signed informed consent. Blood samples were drawn from pregnant mothers, and a blood or buccal sample was collected from the father. Samples from 2 pregnancies with T13 (Patau Syndrome), 2 with T18 (Edwards Syndrome), 2 with T21 (Down's Syndrome), 2 with 45,X, 2 with 47,XXY, and 90 normal pregnancies were selected prior to testing from a cohort of ~500 women to test which chromosomal abnormalities the method detects. Normal fetal karyotype was confirmed by molecular karyotyping for the samples where post-birth child tissue was available. Euploid sample were drawn prior to invasive testing from low-risk women. Aneuploid samples were drawn at least 7 days after invasive testing and aneuploidy was confirmed via cytogenetic karyotyping or fluorescence in situ hybridization at independent laboratories

Sample Preparation and Multiplex PCR

For the data in FIGS. **30**, **30**D, **30**E, **30**G, and **30**H, and **31**A-**31**G, sample preparation and 19,488-plex-PCR were performed as described in Example 15. For the data in FIG. **30**F, sample preparation and 11,000-plex-PCR were performed as described in Example 17.

# US 11,519,035 B2

223

Methodology and Data Analysis

The algorithm considers parental genotypes and crossover frequency data (such as data from the HapMap database) to calculate expected allele distributions for 19,488 polymorphic loci for a very large number possible fetal ploidy states, and at various fetal cfDNA fractions. (FIG. 29). Unlike allele ratio based-methods, it also takes into account linkage disequilibrium, and uses non-Gaussian data models to describe the expected distribution of allele measurements at a SNP given observed platform characteristics and amplification biases. It then compares the various predicted allele distributions to the actual allelic distributions as measured in the cfDNA sample (FIG. 29 step C), and calculates the likelihood of each hypothesis (monosomy, disomy, or trisomy, for which there are numerous hypotheses based on the 15 various potential crossovers) based on the sequencing data. The algorithm sums the likelihoods of each individual monosomy, disomy, or trisomy hypotheses (FIG. 29 step D), and calls the ploidy state with the maximum overall likelihood as the copy number and fetal fraction (FIG. 29 step E). 20 Although laboratory researchers were not blinded to sample karyotype, the algorithm called the ploidy states without human intervention and was blind to the truth.

Data Interpretation

Graphical Representations of the Generated Data

To determine the ploidy state of a chromosome of interest, the algorithm considers the distribution of sequence counts from each of two possible alleles at 3,000 to 4,000 SNPs per chromosome. It is important to note that the algorithm makes ploidy calls using an approach that does not lend 30 itself to visualization. Thus, for the purposes of illustration, the data is displayed here in a simplified fashion as ratios of the two most likely alleles, labeled as A and B, so that the relevant trends can be more readily visualized. This simplified illustration does not take into account some of the 3: features of the algorithm. For example, two important aspects of the algorithm that are not possible to illustrate with a method of visualization that displays allele ratios are: 1) the ability to leverage linkage disequilibrium, i.e. the influence that a measurement at one SNP has on the likely 40 identity of a neighboring SNP, and 2) the use of non-Gaussian data models that describe the expected distribution of allele measurements at a SNP given platform characteristics and amplification biases. Also note that the algorithm only considers the two most common alleles at each SNP, 45 ignoring other possible alleles.

The graphical representations in FIGS. 30, 30D-30H include samples for which two, one, or three fetal chromosomes are present. Generally, these indicate euploidy (FIG. **30**) monosomy (FIG. **30**D), and trisomy (FIGS. **30**E-**30**H), 50 respectively. In all plots, each spot represents a single SNP, where the targeted SNPs are plotted sequentially from left to right for one chromosome along the horizontal axes. The vertical axes indicate the number of reads for the A allele as a fraction of the total number of reads for both the A and B 55 alleles for that SNP. Note that the measurements are made on total cfDNA isolated from maternal blood, and the cfDNA includes both maternal and fetal cfDNA; thus, each spot represents the combination of the fetal and maternal DNA contribution for that SNP. Therefore, increasing the propor- 60 tion of maternal cfDNA from 0% to 100% will gradually shift some spots up or down within the plots, depending on the maternal and fetal genotype. This is described in more detail below with the corresponding plots.

If desired to facilitate visualization, the spots may be 65 color-coded according to maternal genotype, as maternal genotype contributes more to the localization of each spot

224

and the majority of trisomies are maternally-inherited; this assists in visualizing ploidy states. Specifically, SNPs for which the maternal genotype is AA may be indicated in red, those for which the maternal genotype is AB may be indicated in green, and those for which the maternal genotype is BB may be indicated in blue.

In all cases, SNPs that are homozygous for the A allele (AA) in both the mother and the fetus are found tightly associated with the upper limit of the plots, as the fraction of A allele reads is high because there should be no B alleles present. Conversely, SNPs that are homozygous for the B allele in both the mother and the fetus are found tightly associated with the lower limit of the plots, as the fraction of A allele reads is low because there should be only B alleles. The spots that are not tightly associated with the upper and lower limits of the plots represent SNPs for which the mother, the fetus, or both are heterozygous; these spots are useful for identifying fetal ploidy, but can also be informative for determining paternal versus maternal inheritance. These spots segregate based on both maternal and fetal genotypes and fetal fraction, and as such the precise position of each individual spot along the y-axis depends on both stoichiometry and fetal fraction. For example, loci where the mother is AA and the fetus is AB are expected to 25 have a different fraction of A allele reads, and thus different positioning along the y-axis, depending on the fetal fraction. Two Chromosomes Present

FIG. 30 depict data that indicate the presence of two chromosomes when the sample is entirely maternal (no fetal cfDNA present, FIG. 30 (0% FF plot)), contains a moderate fetal cfDNA fraction (FIG. 30 (12% FF plot)), or contains a high fetal cfDNA fraction (FIG. 30 (26% FF plot)).

FIG. 30, 0% FF plot, shows data obtained from cfDNA isolated from the blood of a non-pregnant woman. When there is no fetal cfDNA present and the sample contains only maternal cfDNA, the plots represent purely the euploid maternal genotype; the hallmark pattern includes "clusters" of spots: a filled circle cluster tightly associated with the top of the plot (SNPs where the maternal genotype is AA), a filled square cluster tightly associated with the bottom of the plot (SNPs where the maternal genotype is BB), and a single, centered open triangle cluster (SNPs where the maternal genotype is AB).

When fetal cfDNA is present, the location of the spots shifts such that the clusters segregate into discrete "bands". Note that for samples with a fetal fraction of 0%, the groupings of spots are referred to as "clusters" (as in FIG. 30, 0% FF plot), and for all samples with a fetal fraction of >0%, the groupings of spots are referred to as "bands" (as in FIG. 30 (12% FF plot), 30 (26% FF plot), 30D-30J). If the fetal fraction is high enough, these discrete bands will be readily visible. Specifically, FIG. 30 12% and 26% FF plots demonstrate the characteristic pattern associated with two fetal chromosomes present at moderate and high fetal fractions, respectively. This pattern includes three central open triangle bands that correspond to SNPs that are heterozygous in the mother, and two "peripheral" bands each at both the top (filled circles) and bottom (filled square) of the plots that correspond to SNPs that are homozygous in the mother.

FIG. 30, 12% FF plot, shows data obtained from cfDNA isolated from a plasma sample from a woman carrying a euploid fetus and with a 12% fetal cfDNA fraction. Here, the clusters of spots tightly associated with the top and bottom of the plot segregate into two discrete bands each: one filled circles and one filled square external peripheral band that remains tightly associated with the upper or lower limit of the plots, and one filled circle and one filled square internal

# US 11,519,035 B2

225

peripheral band that has separated from the limits of the plots. These internal peripheral bands, centered around 0.92 and 0.08, represent SNPs for which the maternal genotype is AA and the fetal genotype is AB (indicated in filled circles), and SNPs for which the maternal genotype is BB and the fetal genotype is AB (indicated in filled square), respectively. The center cluster of open triangle spots broadens, but at this fetal fraction the segregation into distinct bands is not readily visible.

At a high fetal cfDNA fraction, the typical pattern that 10 indicates the presence of two chromosomes (a trio of open triangle bands as well as two filled circles and two filled square peripheral bands) is readily apparent. FIG. 30, 26% FF plot, displays data obtained from a plasma sample from a woman carrying a euploid fetus at a fetal cfDNA fraction 15 of 26%. Here, the peripheral bands have separated such that the internal band has shifted towards the center of the plot due to the altered levels of B alleles from the increased fetal cfDNA fraction. Significantly, at higher fetal fractions, the separation of the center open triangle cluster into three 20 distinct bands is now readily apparent. This central trio of bands, in this case clustering around 0.37, 0.50 and 0.63, corresponds to those SNPs where the maternal genotype is AB, and the fetal genotype is AA (top), AB (middle) and BB (bottom)

These hallmark patterns, namely three open triangle bands and four peripheral bands (two filled circles and two filled square), indicate the presence of two chromosomes, as in autosomal euploidy or for the X chromosome in a female (XX) fetus.

### One Chromosome Present

When the fetus only inherits a single chromosome, and thus only inherits a single allele, heterozygosity of the fetus is not possible. As such, the only possible fetal SNP identities are A or B. Thus, maternally-inherited monosomic 35 chromosomes have a characteristic pattern of two central open triangle bands that represent SNPs for which the mother is heterozygous, and only have single peripheral filled circles and filled square bands that represent SNPs for which the mother is homozygous, and which remain tightly 40 associated with the upper and lower limits of the plots (1 and 0), respectively (FIG. 30D). Note the absence of internal peripheral bands. This pattern indicates the presence of one chromosome, as in maternally-inherited autosomal monosomy, or for the X chromosome in a male (XY) fetus.

45 Three Chromosomes Present

Trisomic chromosomes have three characteristic patterns. The first pattern indicates maternally-inherited meiotic trisomy, a meiotic error where the fetus inherited two homologous, non-identical chromosomes from the mother (FIG. 50 30E); this pattern includes two central open triangle bands with two each of the peripheral filled circles and filled square bands. The second pattern indicates paternally-inherited meiotic trisomy, where the fetus inherited two homologous, non-identical chromosomes from the father (FIG. 30F); this 55 pattern includes four central open triangle bands and three each of the peripheral filled circles and filled square bands. The third pattern indicates either maternally—(FIG. 30G) or paternally-inherited (FIG. 30H) mitotic trisomy, a mitotic error where the fetus inherited two identical chromosomes 60 from either the mother or the father; this pattern includes four central open triangle bands with two each of the peripheral filled circles and filled square bands. Maternallyand paternally-inherited mitotic trisomies can be distinguished by the placement of the flanking filled circles and 65 filled square bands, such that the filled circles and filled square internal peripheral bands (those not associated with

226

the limits of the plots) are closer to the center in paternally-inherited mitotic trisomy. This is due to the paternal contribution of identical chromosomes. Note that our previous results indicate that at the blastomere stage, 66.7% of maternally-inherited trisomies are meiotic, and that only 10.2% of trisomies are paternally-inherited.

For the Y chromosome, the PS method considers a different set of hypotheses: zero, one, or two chromosomes present. As there is no maternal contribution to the sequence reads at each locus and because heterozygous loci are not possible (cases of two Y chromosomes necessarily involve two identical chromosomes), the bands remain tightly associated with the top (A alleles) or the bottom (B alleles) of the plot (data not shown), and analysis is greatly simplified, relying on quantitative allele count data. Note that since the method interrogates SNPs, it uses homologous non-recombinant SNPs from the Y chromosome, thus obtaining data on both X and Y for one probe pair.

Identifying Aneuploidies

Identification of autosomal aneuploidies using this plotbased visualization method is straightforward given a sufficient fetal fraction, and requires only identifying plots for which there are an abnormal number of chromosomes present, as described above. Combining the knowledge of copy number of the X and Y chromosomes identifies whether sex chromosome aneuploidies are present. Specifically, plots representing a fetus with a 47,XXX genotype will have a typical "three-chromosome" pattern, and plots representing a fetus with a 47,XXY genotype will have the typical "two-chromosome" pattern for the X chromosome, but will also have allele reads indicating the presence of one Y chromosome. The method is similarly able to call 47,XYY, where a "one chromosome" pattern indicates the presence of a single X chromosome, and allele reads indicate the presence of two Y chromosomes. A fetus with a 45,X genotype will have the typical "one-chromosome" pattern for the X chromosome, and data indicating zero Y chromosomes.

### Effects of Fetal Fraction

As discussed above, the number of sequence reads from the fetus contributes to the precise location of each spot along the y-axis in the plots. As fetal fraction will affect the proportion of reads that originate from the fetus and the mother, it will also affect the positioning of each spot. At a high fraction of fetal cfDNA (generally above ~20%), as in FIGS. 30 (26% FF plot), 30D, 30E, 30G, and 30H, it is readily apparent that although the spots cluster based mainly on maternal genotype, the presence of fetal DNA from alleles whose genotype is distinct from the maternal genotype shift the clusters into multiple, distinct bands. However, as the fetal fraction decreases (as in FIGS. 30 (12% FF plot) and 30F), the spots regress towards the poles and center of the plot, resulting in tighter clusters. Specifically, the set of peripheral filled circles bands, where the maternal genotype is AA, regress towards the top of the plot; the set of peripheral filled square bands, where the maternal genotype is BB, regress towards the bottom; the set of central open triangle bands, where the mother is heterozygous, condense into a single cluster at the center of the plot (compare FIG. 30, 12% and 26% FF plots). Although aneuploidy is not readily apparent by eye using this visualization technique for low fetal fraction cases, the algorithm is able to identify ploidy states with a very low fetal fraction, such as 3% fetal fraction. It is able to do this because the statistical technique compares the observed data to very precise data models that predict the allele distributions for a given sample parameter set (including copy number, parental genotypes, and fetal

# US 11,519,035 B2

227

fraction, for example). Data model precision is critical in low fetal fraction cases, as the differences between the allele distributions for different ploidy states are proportional to the fetal fraction. In addition, the algorithm is able to determine when a data set does not contain enough data to 5 make a confident fetal ploidy determination.

Sequencing reads that mapped to targeted SNPs were deemed to be informative and were used by the algorithm. More than 95% of targeted loci were observed in the sequencing results. The plots for visualizing key ploidy calls are depicted in FIG. 31A-31G. FIG. 31A indicates a euploid sample. Here, chromosomes 13, 18, and 21 have the typical "two chromosome" pattern (as described herein). This includes a trio of center open triangle bands, and two filled circles and two filled square peripheral bands. This, together with the two center open triangle bands for the X chromosome and the presence of Y chromosome bands along the plots' peripheries, indicate a euploid XY genotype.

The most prevalent autosomal trisomies, T13, T18, and T21, are indicated by the plots in FIGS. **31**B, **31**C, and **31**D, respectively. Specifically, FIG. **31**B depicts a T13 sample. Here, chromosomes 18 and 21 display the typical "two chromosome" pattern, chromosome X displays the typical 25 "one chromosome" pattern, and there are reads from the Y chromosome. Together, this indicates disomy at chromosomes 18 and 21, and identifies a fetal XY genotype. However, chromosome 13 depicts a typical "three chromosome" pattern—specifically. Similarly, FIG. **31**C depicts a T18 sample, and FIG. **31**D depicts a T21 sample.

The method is also able to detect sex chromosome aneuploidies, including 45,X (FIG. 31E), 47,XXY (FIG. 31F), and 47,XYY (FIG. 31G). Note that the method is calling copy number at chromosomes 13, 18, 21, X, and Y; the overall chromosome number is reported assuming disomy at the remaining chromosomes. The X chromosome regions of the plot depicting a 45,X sample reveals the presence of a single chromosome. However, the lack of reads from the Y 40 chromosome, coupled with the "two chromosome" pattern for chromosomes 13, 18, and 21, indicate a 45,X genotype. Conversely, the 47,X×Y samples generate a plot revealing the presence of two X chromosomes. The data also revealed reads for alleles from the Y chromosome. Together with the 45 presence of two copies of chromosomes 13, 18, and 21, this indicates a 47,XXY genotype. A 47,XYY genotype is indicated by the presence of a "one chromosome" pattern for the X chromosome, and reads indicating the presence of two Y chromosomes

### Discussion

This method detected T13, T18, T21, 45,X, 47,XXY, and 47,XYY non-invasively from maternal blood. This method interrogates cfDNA from maternal plasma by targeted multiplex PCR amplification and high-throughput sequencing of 55 19,488 SNPs. This, coupled with the method's sophisticated informatics analyses that take into account parental genotypic information and numerous sample parameters, including fetal fraction and DNA quality, more robustly detects the fetal signal and makes highly accurate ploidy calls at all of 60 the five chromosomes implicated in the seven most common types of at-birth aneuploidy (T13, T18, T21, 45,X, 47,XXX, 47,XXY, and 47,XYY). This method offers a number of clinical advantages over previous methods, including and most significantly greater clinical coverage and sample- 65 specific calculated accuracies (analogous to a personalized risk score).

228

Increased Clinical Coverage

This method offers approximately a two-fold increase in aneuploidy coverage compared to clinically available NIPT methodologies, given its ability to accurately detect autosomal trisomies and sex chromosome aneuploidies. The method presented here is the only noninvasive test that calls ploidy at the sex chromosomes with high accuracy. Prior DNA mixing experiments and separate plasma samples analyzed in our experimental assays suggest that this method will detect a larger cohort of sex chromosome anomalies, including 47,XXX. The method presented here also detects aneuploidies at chromosomes 13, 18, and 21 with high sensitivities and specificities, and with appropriate primer design is expected to be able to detect copy number at the remaining chromosomes as well.

5 Sample-Specific Calculated Accuracies

Significantly, this method calculates a sample-specific accuracy for ploidy calls on each chromosome in each sample. Accuracies calculated by this method are expected to significantly lower the rate of incorrect calls by identify-20 ing and flagging individual samples that have poor quality DNA or low fetal fractions that are likely to result in a poor accuracy test result. By contrast, massively parallel shotgun sequencing (MPSS)-based methods produce a positive or negative call using a single-hypothesis rejection test, and their accuracy estimate is based on a published study cohort rather than on the characteristics of the individual sample, which are assumed to have the same accuracy as the cohort. However, individual accuracies for samples with parameters in the tail of the cohort distribution may differ significantly. This is exacerbated at low fetal fractions, as in early gestational age, or for samples with low DNA quality. These samples are generally not identified and flagged for followup, which can result in missed calls. The present method, however, takes into account many parameters, including fetal fraction and a number of DNA quality metrics, to make each chromosome copy number call, calculating a samplespecific accuracy for that call. This allows the method to identify individual samples with low accuracy and flag them for follow-up. This is expected to nearly eliminate missed calls, especially at the early stages of pregnancy when fetal fractions are typically low. The presumption is that a no call is much preferred to a missed call, since a no call simply requires a redraw and reanalysis.

Converting Calculated Accuracies to Traditional Risk Scores

This method can offer an adjusted risk of aneuploidy for high-risk pregnant women, where the adjusted risk takes into account an a priori risk (Benn P, Cuckle H, Pergament E. Non-invasive prenatal diagnosis for Down syndrome: the paradigm will shift, but slowly. Ultrasound Obstet Gynecol 2012; 39:127-130, which is hereby incorporated by reference in its entirety). Although the present method offers each patient a customized calculated accuracy, for clinical use these accuracies can be converted to traditional risk scores, which also denote the risk of an aneuploid pregnancy but are expressed as fractions. Traditional risk scores take into account various parameters, including maternal age-related risk and serum levels of biochemical markers, to offer a risk score above which a mother is considered high-risk and for whom follow-up invasive diagnostic procedures are recommended. This method significantly refines this risk score, thus reducing both the false positive and false negative rates, and offering a more accurate assessment of individual maternal risk. A calculated accuracy as used here is the likelihood that the ploidy call is correct, and is expressed as a percentage, but the calculated accuracies used in Experiment 19 do not include an age-related risk. Because calculation of a risk

# US 11,519,035 B2

229

score typically includes an age-related risk, the calculated accuracies and traditional risk scores are not interchangeable; they must be combined to convert into a traditional risk score. The formula to combine the age-related risk with the calculated accuracy is:

$$\frac{R_1R_2}{R_1R_2 + [1-R_1][1-R_2]}$$

where  $R_1$  is the risk score as calculated by the present method and  $R_2$  is the risk score as calculated by first trimester screening.

SNP-Based Methods Negate Issues with Amplification <sub>15</sub> Variation

An inherent drawback to the counting methods used by some other methods is that they determine fetal ploidy state by measuring the ratio of the number of reads mapping to the chromosome of interest (e.g., chromosome 21) to those 20 mapping to a reference chromosome. Chromosomes with high or low GC content, including chromosomes 13, X, and Y, amplify with high variability. This can result in signal variation that is comparable in magnitude to the fetal cfDNA signal, which can confound copy number calls by altering the ratio of allele reads from the chromosome-of-interest to those from the reference chromosome. This can result in low accuracy for chromosomes 13, X, and Y. Significantly, this problem is exacerbated at low fetal cfDNA fractions, as tends to be the case at early gestational ages.

In contrast, SNP-based methods do not rely on consistent amplification levels between chromosomes, and are thus expected to provide results that are equally accurate across all chromosomes. Because the present method looks, in part, at relative counts of different alleles at polymorphic loci, which by definition differ only by a single nucleotide, it does not require the use of reference chromosomes, and this obviates the problems with chromosome-to-chromosome amplification variation that are inherent to methods that rely on quantitating read counts. Unlike quantitative methods that require reference chromosomes that are euploid, the present method is expected to be able to detect triploidy as well as copy-number neutral anomalies like uniparental disomy.

The Importance of Early Detection

Significantly, the combined at-birth prevalence of sex chromosome aneuploidies is higher than that of the most common autosomal aneuploidies (FIG. 32). However, there are currently no routine non-invasive screening methods that 50 reliably detect sex chromosome abnormalities. Thus, sex chromosome anomalies are generally detected prenatally as a side-effect of routine testing for Down syndrome or other autosomal aneuploidies; a large proportion of cases are missed entirely. Early and accurate detection is crucial for 55 many of these disorders where early therapeutic intervention improves clinical outcomes. For example, Turner syndrome is often not diagnosed until adolescence, although its overall at-birth prevalence is 1 in 2,500 females. Growth hormone therapy is known to prevent short stature that results from 60 the disorder, but treatments are significantly more effective when initiated prior to the age of 4. Additionally, estrogen replacement therapy can stimulate secondary sexual characteristics in patients with Turner syndrome, but again therapy must be initiated in the pre-teen years, before the 65 syndrome is usually detected. Together, this underscores the importance of early, routine, and safe detection of sex

230

chromosome aneuploidies. This method offers the first approach with the potential to serve as a routine screen for sex chromosome anomalies.

### Example 20

The following experiment illustrates an exemplary method for designing and selecting primers that can be used in any of the multiplexed PCR methods of the invention. In some embodiments, primers from an initial library of candidate primers are selected so that they can be used to simultaneously amplify a large number of target loci (or a subset of target loci) in a single reaction. In some embodiments, primers from an initial library of candidate primers are selected to form multiple primer pools such that each pool can be used to simultaneously amplify a subset of target loci in a single reaction. Preferably, primers are designed and selected for a large portion or all of the most desirable target loci. Preferably, the minimum number of pools needed to amplify the target loci are created.

Step 1

Calculate a first score for each primer pair design using one or more of the following parameters: number of SNPs within the primers, location of SNPs within the primers, distance from an end of the amplicon to the target bases within the amplicon, number of target loci in an amplicon, heterozygosity rate of the target locus, disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, specificity of the candidate primer for the target locus, size of the candidate primer, melting temperature of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, GC content of the 3' end of the candidate primer, homopolymer length in the candidate primer, amplification efficiency of the target amplicon, and size of the target amplicon. Step 2

Compare each primer pair to every other primer pair, and calculate a second score for the pair using one or more of the following parameters: likelihood of dimer formation, amplicon overlap, number of primer designs for a particular target locus, and distance between amplicons. In some embodiments, the score is infinite if amplicons overlap so that two different primer pairs that generate overlapping amplicons are not included in the same primer pool.

Step 3

Aggregate the first score and the second score together (such as by using a weighted average of the scores). Step 4

If desired, order all target loci into one contiguous list based upon their genomic location in ascending order.

Step 5

Build a minimum priority queue data structure that prioritizes the pairs of designs (in which each design is one primer pair so that a pair of designs includes two primer pairs with a total of 4 primers) based on their score (such as the aggregate score from step 3). In some embodiments, the score for a pair of designs is the worse score (such as the worse aggregate score from step 3) out of the scores for all 4 primers in the pair of designs. The pair of designs with the best (most desirable) score is first in the queue, and the pair of designs with the worst (least desirable) score is last in the queue. If desired, pairs of designs with a score above a threshold (least desirable) are removed from the library of candidate primers such that they are not included in the final pool(s) (for example, these primers may be omitted from the

# US 11,519,035 B2

23

queue). In some embodiments, pairs of design with an interaction score above (worse than) 20, 18, 16, 14, 12, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 kcal/mol are removed from the library of candidate primers. In some embodiments, pairs of design with a  $\Delta G$  value below (worse than) -20, -18, -16, -14, -12, -10, -9, -8, -7, -6, -5, -4, -3, -2, or -1 kcal/mol are removed from the library of candidate primers.

Each design pair can be stored as a node of a doubly linked list with initial "next" and "previous" pointers set to NIII I

Step 6

Begin forming all pools simultaneously by doing the following steps. Take the design pair with the best (most desirable) score from the priority queue and add it to "the potential pools." Begin storing designs in N number of 15 doubly linked list data structures with the design pairs. N represents the current number of different primer pools. Initially, N=1, since there is only one primer pool. In some embodiments, a second pool is only created if necessary to include the desired target loci or the desired level of cov- 20 erage of target loci. Check to see if the design pair removed from the queue is "connected" to any other existing design pair. By "connected" for purposes of this step is meant that a single design in one pair is the same as a single design in another pair. If two pairs are connected, then assign the 25 appropriate next and previous pointers to one another. If two pairs are not connected, then add them to the "potential pools" In some embodiments, a design pair is only placed in a particular pool if it would be connected to at most two other design pairs in that pool (otherwise it can be assigned 30 to a different pool).

Check to see if (i) any linked list spans from the first target to the last target (such that all the desired target loci are included) or (ii) if a pool meets the cutoff for the desired minimum pool level. If it does, that list now forms a pool and 35 can be added to the "final pools" list.

Step 7

If desired, check to see if the desired level of coverage (such as all the bases in the target loci being included in amplicons from 4 different primer pairs) that is desired for 40 each location. Repeat step 6 until achieving the desired level of coverage.

The resulting primer pool(s) can be used in any of the methods of the invention.

# Example 21

The following Example illustrates an exemplary method for designing and selecting primers that can be used in any of the multiplexed PCR methods of the invention. In some 50 embodiments, the primers are divided into different pools (e.g., 2, 3, 4, 5, 6, or more different pools) such that each pool is used to amplify target loci in a different reaction volume. Each pool is used to simultaneously amplify a large number of target loci (or a subset of target loci) in a single 55 reaction volume. Preferably, primers are designed and selected for a large portion of the most desirable target loci or for all of the target loci. A set of candidate target loci can be selected as described in Examples 16 or 20 based on the particular polymorphisms or mutations of interest. In some 60 embodiments, one or more of the following type of target loci are included: SNPs, short indels, long indels, exons, and combinations thereof. In some embodiments for target loci that are short indels, the PCR primer or primer pair targets a sequence of adjacent base pairs; and the indel is com- 65 pletely covered by one sequencing read. In some embodiments for target loci that are large indels, two primer pairs

232

are used to target a pair of breakpoints at the boundaries of the indel. In this case, the two primer pairs are designed such that when the deletion is present there is a PCR product and the two primer pairs are selected together for inclusion in the same pool (the four primers are treated by the algorithm as a single assay rather than two assays). In some embodiments for target loci that are exons, a set of primers pairs are designed to tile the full exon.

For each candidate locus, one or more PCR primer pairs are designed using the Primer3 program (available at the worldwide web at primer3.sourceforge.net; libprimer3 release 2.2.3, which is hereby incorporated by reference in its entirety). If there are no feasible designs for PCR primers for a particular target locus, then that target locus is eliminated from further consideration. In some embodiments, each target base is covered by at least two independent PCR assays (such as two independent primer pairs that will amplify the target base) and preferably by four assays, although not all of the available assays for a target must be used. In some embodiments, no targets are omitted. Desirably, the algorithm produces as few pools as possible but may produce more than one pool. In some embodiments, two different primer pairs that are in close proximity in the genome (such as within 2 kbases or 1 kbase) and whose forward primers are on the same strand are not be assigned to the same pool. This constraint avoids primer interference in the extension-and-ligation amplification method In some embodiments in which the PCR will be performed using a polymerase with low 5'→3' exonuclease and/or low strand displacement activity, different primer pairs that are in close proximity in the genome and whose forward primers are on the same strand can be assigned to the same pool since the with low 5'→3' exonuclease and/or low strand displacement activity of the polymerase will reduce or prevent primer interference and allow nearby or adjacent amplicons to be produced.

Step 1

Build an interaction graph. Each node represents one assay (such as one primer pair). Each edge represents a conflict between two assays. There are three types. Interaction edges represent a potential primer dimer and have a score indicating the interaction strength. Proximity edges represent physical proximity of the primer binding sites which may result in interference. Target edges represent redundant designs associated with the same target (a special case of a proximity edge).

Step 2

Select an initial value for the maximum interaction score (e.g., 95% of the maximum score).

Step 3

Compute a score such as a utility score for each assay as follows using steps 3A and 3B.

Step 3A

Step 3A

Calculate a score for each assay based on one or more of its intrinsic characteristics. For example, favor assays with amplicons close to the optimal length (such as 300 bp); favor assays with a shorter distance from the beginning of the amplicon to the target; and/or penalize assays with primers overlapping known SNPs. Any other parameter, such as the parameters disclosed herein can also be included. Step 3B

Multiply the score for each assay by a factor that varies from 0 to 1 according to the current coverage of the assay's target bases. This factor gives lower weight to targets that are already covered by assays. At the beginning of the algorithm this factor is 1 for all assays because none have been covered. Calculate the factor as follows. For each base

# US 11,519,035 B2

### 233

in the target, compute a coverage score as  $1/(2^{\circ}c)$  where c is the number of previously-selected assays (in other pools) that cover that base. For instance, if three assays cover the base then the coverage score is  $1/(2^{4}3)=0.125$ . The factor for the target is the maximum value of the coverage score for all bases in the target. For instance, if the target contains 10 bases, 3 bases are covered by 1 target, and 7 bases are covered by 3 targets, then the factor is  $MAX(1/(2^{4}1), 1/(2^{4}3))=0.5$ . The score in step 3A is then multiplied by this factor.

Step 4

Use a single iteration of the algorithm in Example 16 to design a pool given the current maximum interaction score: Construct a new graph with the assays that have not been assigned to a pool yet and with the edges that have weights exceeding the maximum interaction score. Remove nodes (assays) according to the algorithm in Example 16 until there are no edges left. The assay utility scores come from step 3 in this Example rather than the calculation used for 20 Example 16.

Step 5

Save the assays selected in step 4 as a new pool and remove them from consideration. Then repeat steps 3 and 4 with the remaining assays, and iterate until all targets have 25 sufficient coverage.

Step 6

If desired, evaluate the result. If the total number of pools meets the design goal then reduce the maximum interaction score; otherwise increase the maximum interaction score. <sup>30</sup> Then go back to step 3. Iterate, using a binary search strategy to find the lowest maximum interaction score that produces the desired number of pools.

Step 7

Output the pools from the final iteration. After the selection process, the primers remaining in the pools may be used in any of the methods of the invention.

### Example 22

The following Example illustrates an exemplary method for designing and selecting primers that can be used in any of the multiplexed PCR methods of the invention. In some embodiments, the primers are divided into different pools (e.g., 2, 3, 4, 5, 6, or more different pools) such that each 45 pool is used to amplify target loci in a different reaction volume. Any of the embodiments listed in Example 21 can be used for this Example as well.

This method uses a graph coloring algorithm. Step 1

Select 2, 3 or 4 of the best assays (such as primer pairs) for each target locus from all of the available assays. Step 2

Select an initial maximum interaction score. Step 3

Build an interaction graph containing only edges that exceed the maximum interaction score.

Step 4

Color the graph such that no adjacent nodes have the same color (this is a standard problem with many heuristic solutions). Each color represents a different pool.

Step 5

Go back to step 3 and iterate, refining the maximum interaction score until the desired number of pools is achieved. In some embodiments, after the primers are 65 selected in step 1, the algorithm assumes all assays must be included in a pool.

### 234

After the primers are divided into different pools, the pools may be used in any of the methods of the invention.

### Example 23

This example illustrates there exemplary methods for calculating the limit of detection for any of the methods of the invention. These methods were used to calculate the limit of detection for single nucleotide variants (SNVs) in a tumor biopsy (FIG. 38) and a plasma sample (FIG. 39).

The first method (denoted "LOD-mr5" in FIGS. 38 and 39) calculates the limit of detection based on a minimum of 5 reads being chosen as the minimum number of times a SNV is observed in the sequencing data to have sufficient confidence the SNV is actually present. The limit of detection is based on whether the observed the depth of read (DOR) is above this minimum of 5. The thin lines (LOD-z5.0) in FIGS. 38 and 39 indicate SNVs for which the limit of detection is limited by the DOR. In these cases, not enough reads were measured to reach the error limit of the assay. If desired, the limit of detection can be improved (resulting in a lower numerical value) for these SNVs by increasing the DOR.

The second method (denoted "LOD-zs5.0" in FIGS. 38 and 39) calculates the limit of detection based on the z-score. The Z-score is the number of standard deviations an observed error percentage is away from the background mean error. If desired, outliers can be removed and the z-score can be recalculated and this process can be repeated. The final weighted mean and the standard deviation of the error rate are used to calculate the z-score. The mean is weighted by the DOR since the accuracy is higher when the DOR is higher.

For the exemplary z-score calculation used for this example, the background mean error and standard deviation were calculated from all the other samples of the same sequencing run weighted by their depth of read, for each genomic loci and substitution type. Samples were not considered in the background distribution if they were 5 standard deviations away from the background mean. The dashed lines in FIGS. 38 and 39 indicate SNVs for which the limit of detection is limited by the error rate. For these SNV's enough reads were taken to reach the 5 read minimum, and the limit of detection was limited by the error rate. If desired, the limit of detection can be improved by optimizing the assay to reduce the error rate.

The third method (denoted "LOD-zs5.0-mr5" in FIGS. 38 and 39) calculates the limit of detection based on the maximum value of the above two metrics.

For the analysis of a tumor sample shown in FIG. 38, the mean limit of detection was 0.36%, and the median limit of detection was 0.28%. The number of DOR limited (thin lines) SNVs was 934. The number of error rate limited (dashed lines) SNVs was 738.

For the analysis of cDNA in a plasma sample shown in FIG. **39**, the mean limit of detection was 0.24%, and the median limit of detection was 0.09%. The number of DOR limited (thin lines) SNVs was 732. The number of error rate limited (dashed lines) SNVs was 921.

# Example 24

This example illustrates the detection of CNVs and SNVs from the same single cell. The following primer libraries were used a library of ~28,000 primers for detecting CNVs, a library of ~3,000 primers for detecting CNVs, and library of primers for detecting SNVs. For analysis of a single cell,

# US 11,519,035 B2

235

cells were serial diluted until there were 3 or 4 cells per droplet. An individual cell was pipetted and placed into a PCR tube. The cell was lysed using Protease K, salt, and DTT using the following thermocycling conditions: 56° C. for 20 minutes, 95° C. for 10 minutes, and then a 4° C. hold. 5 For analysis of genomic DNA, DNA from the same cell line as the analyzed single cell was either purchased or obtained by growing the cells and extracting the DNA.

For amplification with the library of ~28,000 primers, the following PCR conditions were used a 40 uL reaction 10 volume, 7.5 nM of each primer, and 2× master mix (MM). In some embodiments QIAGEN Multiplex PCR Kit is used for the master mix (QIAGEN catalog No. 206143; see, e.g., information available at the world wide web at giagen.com/ products/catalog/assay-technologies/end-point-pcr-and-rtpcr-reagents/qiagen-multiplex-pcr-kit, is which is hereby incorporated by reference in its entirety). The kit includes 2× QIAGEN Multiplex PCR Master Mix (providing a final concentration of 3 mM MgCl<sub>2</sub>, 3×0.85 ml), 5× Q-Solution (1×2.0 ml), and RNase-Free Water (2×1.7 ml). The QIA- 20 GEN Multiplex PCR Master Mix (MM) contains a combination of KCl and (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as well as the PCR additive, Factor MP, which increases the local concentration of primers at the template. Factor MP stabilizes specifically bound primers, allowing efficient primer extension by, e.g., Hot- 25 StarTaq DNA Polymerase. HotStarTaq DNA Polymerase is a modified form of Taq DNA polymerase and has no polymerase activity at ambient temperatures. The following thermocycling conditions were used for the first round of PCR: 95° C. for 10 minutes; 25 cycles of 96° C. for 30 30 seconds, 65° C. for 29 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and a 4° C. hold. For the second round of PCR a 10 ul reaction volume, 1×MM, and 5 nM of each primer was used. The following thermocycling conditions were used: 95° C. for 15 minutes; 25 cycles of 94° C. for 30 seconds, 65° C. for 1 minute, 60° C. for 5 minutes, 65° C. for 5 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and a 4° C. hold.

For the library of ~3,000 primers, exemplary reaction conditions include a 10 ul reaction volume, 2× MM, 70 mM 40 TMAC, and 2 nM primer of each primer. For the library of primers for detecting SNVs, exemplary reaction conditions include a 10 ul reaction volume, 2× MM, 4 mM EDTA, and 7.5 nM primer of each primer. Exemplary thermocycling conditions include 95° C. for 15 minutes, 20 cycles of 94° 45 C. for 30 seconds, 65° C. for 15 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and a 4° C. hold.

The amplified products were barcoded. One run of sequencing was performed with an approximately equal number of reads per sample.

FIGS. **40**A and **40**B show results from analysis of genomic DNA (FIG. **40**A) or DNA from a single cell (FIG. **40**B) using a library of approximately 28,000 primers designed to detect CNVs. Approximately 4 million reads were measured per sample. The presence of two central bands instead of one central band indicates the presence of a CNV. For three samples of DNA from a single cell, the percent of mapped reads was 89.9%, 94.0%, and 93.4%, respectively. For two samples of genomic DNA the percent of mapped reads was 99.1% for each sample.

FIGS. **41**A and **41**B show results from analysis of genomic DNA (FIG. **41**A) or DNA from a single cell (FIG. **41**B) using a library of approximately 3,000 primers designed to detect CNVs. Approximately 1.2 million reads were measured per sample. The presence of two central 65 bands instead of one central band indicates the presence of a CNV. For three samples of DNA from a single cell, the

236

percent of mapped reads was 98.2%, 98.2%, and 97.9%, respectively. For two samples of genomic DNA the percent of mapped reads was 98.8% for each sample. FIG. 42 illustrates the uniformity in DOR for these ~3,000 loci.

For calling SNVs, the call percent for true positive mutations was similar for DNA from a single cell and genomic DNA. A graph of call percent for true positive mutations for single cells on the y-axis versus that for genomic DNA on the x-axis yielded a curve fit of y=1.0076x-0.3088 with R<sup>2</sup>=0.9834. FIG. 43 shows similar error call metrics for genomic DNA and DNA from a single cell. FIG. 44 shows that the error rate for detecting transition mutations was greater than for detecting transversion mutations, indicating it may be desirable to select transversion mutations for detection rather than transition mutations when possible. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 98, 99, or 100% of the SNVs tested for are transversion mutations rather than transition mutations.

### Example 25

The following is an example of multiplex PCR conditions in which the annealing temperature is significantly higher than the average or maximum melting  $(T_m)$  of the primers in the library. A 3,168-plex reaction was performed with 3,168 primer pairs to 3,168 different target loci. For the PCR amplification a 20 ul total volume was used with 2 nM of each primer (3,168 pairs of forward and reverse primers), 70 mM TMAC (tetra-methyl ammonium chloride), and 7 ul library DNA or genomic DNA. The following thermocycling conditions were used: 95° C. for 10 minutes and then 25 cycles of 96° C. for 30 seconds, 65° C. for 20 minutes (this annealing temperature is higher than the  $T_m$  of the primers, listed above), and 72° C. for 30 seconds. Then, 72° C. for 2 minutes and a 4° C. hold were used.

The minimum  $T_m$  (the lowest numerical value for the  $T_m$  for any of the primers) for this primer library is  $54.0^{\circ}$  C. The maximum  $T_m$  (the highest numerical value for the  $T_m$  for any of the primers) for this primer library is  $60.36^{\circ}$  C. The average  $T_m$  (average value of the  $T_m$  values of the primers) for this primer library is  $55.25^{\circ}$  C. These  $T_m$  values were calculated using the following exemplary method for calculating  $T_m$  values. This method is used by the Primer3 program (the worldwide web at primer3.sourceforge.net, which is hereby incorporated by reference in its entirety) to calculate  $T_m$  values. In some embodiments, one or more of the following conditions are assumed for this calculation temperature: of  $60.0^{\circ}$  C., primer concentration of 100 nM. In some embodiments, other conditions are assumed for this calculation, such as the conditions that will be used for multiplex PCR with the library.

### $T_m \!\!=\!\! \mathrm{delta} H \! / \! (\mathrm{delta} S \!\!+\!\! R^* \! \ln(C/\!4))$

Below is documentation from the Primer3 program for its  $T_m$  calculations; PRIMER\_TM\_FORMULA (int; default 0) specifies details of melting temperature calculation. This is new in version 1.1.0, and added by Maido Remm and Triinu Koressaar (the world wide web at primer3.ut.ee/primer3web\_help.htm #PRIMER\_TM\_FORMULA, which is hereby incorporated by reference in its entirety). A value of 0 directs primer3 to a backward compatible calculation (in other words, the only calculation available in previous version of primer3). This backward compatible calculation uses the table of thermodynamic parameters in the paper (Breslauer K J et al. (1986) "Predicting DNA duplex stability from the base sequence" Proc Natl Acad Sci 83:4746-50,

# US 11,519,035 B2

dx.doi.org/10.1073/pnas.83.11.3746, which is hereby incorporated by reference in its entirety), and the method in the paper (Rychlik W, Spencer W J and Rhoads R E (1990) 'Optimization of the annealing temperature for DNA amplification in vitro", Nucleic Acids Res 18:6409-12, dx.doi.org/ 10.1093/nar/18.21.6409, which is hereby incorporated by reference in its entirety).

A value of 1 (which is recommended) directs primer3 to use the table of thermodynamic values and the method for melting temperature calculation suggested in the following 10 paper (SantaLucia JR (1998) "A unified view of polymer, dumbbell and oligonucleotide DNA nearest-neighbor thermodynamics", Proc Natl Acad Sci 95:1460-65, dx.doi.org/ 10.1073/pnas.95.4.1460, which is hereby incorporated by reference in its entirety). The tag PRIMER\_SALT\_COR- 15 RECTIONS can be used to specify the salt correction method for melting temperature calculation.

The following is an example of calculating the melting temperature of an oligo with PRIMER\_TM\_FORMULA=1 and PRIMER\_SALT\_CORRECTIONS=1 recommended 20 values for primer=CGTGACGTGACGGACT.

Using default salt and DNA concentrations gives

 $T_m$ =deltaH/(deltaS+R\*ln(C/4))

DNA concentration.

$$deltaH(predicted) = dH(CG) + dH(GT) + dH(TG) + ... + dH(CT) + \\ dH(init.w.term.GC) + dH(init.w.term.AT)$$
 
$$= -10.6 + (-8.4) + (-8.5) + ... + (-7.8) + 0.1 + 2.3$$
 
$$= -128.8 \text{ kcal/mol}$$

where 'init.w.term GC' and 'init.w.term AT' are two initiation parameters for duplex formation: 'initiation with terminal GC' and 'initiation with terminal AT.'

$$\begin{aligned} deltaS(\text{predicted}) &= dS(CG) + dS(GT) + dS(TG) + \dots + dS(CT) + \\ &= dS(init.w.term.GC) + dS(init.w.term.AT) \\ &= -27.2 + (-22.4) + (-22.7) + \dots + (-21.0) + \\ &= (-2.8) + 4.1 \\ &= -345.2 \text{ cal/k} * \text{mol} \end{aligned}$$
 
$$\begin{aligned} deltaS(\text{salt corrected}) &= deltaS(\text{predicted}) + 0.368 * 15(NN \text{ pairs}) * \\ &= \ln(0.05M \text{ monovalent cations}) \end{aligned}$$
 
$$= -361.736$$
 
$$Tm = -128.800/(-361.736 + 1.987 * \ln((5*10^{\circ}(-8))/4)) = 323.704 \text{ K}$$
 
$$Tm(C) = 323.704 - 273.15 = 50.554 \text{ C}$$

# Additional Applications

Because this method utilizes targeted amplification, it is uniquely poised to detect submicroscopic anomalies, such as microdeletions and microduplications. Although non-targeted methods like MPSS have been shown to detect the 60 DiGeorge microdeletion syndrome, this required a sufficiently high level of genomic coverage so as to make the approach unfeasible. This is because non-targeted amplification will be several orders of magnitude less efficient on submicroscopic regions, as very small fraction of the 65 sequencing reads will be informative. Additionally, the fact that the currently available methods have trouble accurately

### 238

identifying ploidy state for the sex chromosomes suggests that they will also encounter variable amplification problems on smaller chromosomal segments.

Similarly, SNP based methods can detect UPD disorders, which are copy number-neutral anomalies that will not be detected by either the current noninvasive methods that rely on counting or the traditional invasive methods like amniocentesis and CVS that rely on cytogenetic karyotyping and/or fluorescence in situ hybridization. This is because SNP-based methods are uniquely able to distinguish individual haplotypes, whereas the clinically available MPSSbased and targeted methods amplify non-polymorphic loci and are thus unable to determine, for example, whether the chromosomes-of-interest originate from the same parent. This means that these microdeletion/microduplication and UPD syndromes, including Prader-Willi, Angelman, and Beckwith-Wiedemann syndromes, are generally not diagnosed prenatally, and are often initially misdiagnosed postnatally. This significantly delays therapeutic intervention. Additionally, because this method targets SNPs, this method will also facilitate parental haplotype reconstruction, allowing for detection of fetal inheritance of individual diseaselinked loci (Kitzman J O, Snyder M W, Ventura M, et al. Noninvasive whole-genome sequencing of a human fetus. where R is the gas constant (1.987 cal/K mol) and C is the 25 Sci Transl Med 2012; 4:137ra76, which is hereby incorporated by reference in its entirety).

The results presented here confirm the expanded scope of this method for identifying prenatal aneuploidy. Specifically, by amplifying and sequencing 19,488 SNPs, this method is 30 able to determine copy number at chromosomes 13, 18, 21, X, and Y, and is uniquely expected to detect other chromosomal abnormalities, such as triploidy and UPD, that are not detected by any other clinically available non-invasive method. The increased clinical coverage and powerful sample-specific calculated accuracies suggest that this method may offer a viable adjunct to invasive testing for detecting fetal chromosomal aneuploidies.

# Example 26

40 This example describes an exemplary method for detection of copy number variations in breast cancer samples using SNP-targeted massively multiplexed PCR. Evaluation of CNV in tumor tissues typically involves SNP microarray 45 or aCGH. These methods have high whole-genome resolution, but require large amounts of input material, have high fixed costs, and do not work well on formaldehyde fixedparaffin embedded (FFPE) samples. For this example, 28,000-plex SNP-targeted PCR with next generation 50 sequencing (NGS) was used to target ip, 1q, 2p, 2q, 4p16, 5p15, 7q11, 15q, 17p, 2211, 2213 and chromosomes 13, 18, 21 and X for detection of CNVs in breast cancer samples. Accuracy was validated on 96 samples with aneuploidies or microdeletions. Single-molecule sensitivity was established by analyzing single cells. Of 17 breast cancer samples (15 fresh frozen and 2 FFPE tumor tissues, 5 pairs of matched tumor and normal cell lines) analyzed, 16 (including both FFPEs) were observed with full or partial CNVs in one to 15 targets (average: 7.8); evidence of tumor heterogeneity was observed. The three tissues with one CNV all had a 1q duplication, the most frequent cytogenetic abnormality in breast carcinoma. The most frequent regions with CNVs were 1q, 7p, and 221. Only one tumor tissue (with 9 CNVs) had a region with LOH; this LOH was also detected in adjacent putatively normal tissue that lacked the other 8 CNVs. By contrast, 5 or more regions with LOH and a high total CNV incidence (average: 12.8) was detected in cell

# US 11,519,035 B2

### 239

lines. Thus, massively multiplexed PCR offers an economical high-throughput approach to investigate CNVs in a targeted manner, and is applicable to difficult-to-analyze samples, such as FFPE tissues.

### Example 27

This example further validates a massively multiplexed PCR methodology for chromosomal aneuploidy and CNV determination disclosed herein, sometimes referred to as 10 CoNVERGe (Copy Number Variant Events Revealed Genotypically), in cancer diagnostics, and further illustrates the development and use of "PlasmArt" standards for PCR of ctDNA samples. PlasmArt standards include polynucleotides having sequence identity to regions of the genome 15 known to exhibit CNV and a size distribution that reflects that of cfDNA fragments naturally found in plasma. Sample Collection

Human breast cancer cell lines (HCC38, HCC1143, HCC1395, HCC1937, HCC1954, and HCC2218) and 20 matched normal cell lines (HCC38BL, HCC1143BL, HCC1395BL, HCC1937BL, HCC1954BL, and HCC2218BL) were obtained from the American Type Culture Collection (ATCC). Trisomy 21 B-lymphocyte (ΔG16777) and paired father/child DiGeorge Syndrome 25 (DGS) cell lines (GM10383 and GM10382, respectively) were from the Coriell Cell Repository (Camden, N.J.). GM10382 cells only have the paternal 2211.2 region.

We procured tumour tissues from 16 breast cancer patients, including 11 fresh frozen (FF) samples from 30 Geneticist (Glendale, Calif.) and five formalin-fixed paraffin-embedded (FFPE) samples from North Shore-LIJ (Manhasset, N.Y.). We acquired matched buffy coat samples for eight patients and matched plasma samples for nine patients. FF tumour tissues and matched buffy coat and plasma 35 samples from five ovarian cancer patients were from North Shore-LIJ. For eight breast tumour FF samples, tissue subsections were resected for analysis. Institutional review board approvals from Northshore/LIJ IRB and Kharkiv National Medical University Ethics Committee were 40 obtained for sample collection and informed consent was obtained from all subjects.

Blood samples were collected into EDTA tubes. Circulating cell free DNA (containing ctDNA) was isolated from 1 mL plasma using the QIAamp Circulating Nucleic Acid 45 Kit (Qiagen, Valencia, Calif.). Genomic DNA (gDNA) from FF tumor tissues, blood, and buccal samples was extracted using the DNeasy Blood and Tissue Kit (Qiagen).

To make the PlasmArt standards according to one exemplary method, first, 9×10<sup>6</sup> cells were lysed with hypotonic 50 lysis buffer (20 mM Tris-Cl (pH 7.5), 10 mM NaCl, and 3 mM MgCl<sub>2</sub>) for 15 min on ice. Then, 10% IGEPAL CA-630 (Sigma, St. Louis, Mo.) was added to a final concentration of 0.5%. After centrifugation at 3,000 g for 10 min at 4° C., pelleted nuclei were resuspended in 1× micrococcal nucle- 55 ase (MNase) Buffer (New England BioLabs, Ipswich, Mass.) before adding 1000 U of MNase (New England BioLabs), and then incubated for 5 min at 37° C. Reactions were stopped by adding EDTA to a final concentration of 15 mM. Undigested chromatin was removed by centrifugation 60 at 2,000 g for 1 min. Fragmented DNA was purified with the DNA Clean & Concentrator<sup>TM</sup>\_500 kit (Zymo Research, Irvine, Calif.). Mononucleosomal DNA produced by MNase digestion was also purified and size-selected using AMPure XP magnetic beads (Beckman Coulter, Brea, Calif.). DNA 65 fragments were sized and quantified with a Bioanalyzer DNA 1000 chip (Agilent, Santa Clara, Calif.).

### 240

To model ctDNA at different concentrations, different fractions of PlasmArts from HCC1954 and HCC2218 cancer cells were mixed with those from the corresponding matched normal cell line (HCC1954BL and HCC2218BL, respectively). Three samples at each concentration were analyzed. Similarly, to model allelic imbalances in plasma DNA in a focal 3.5 Mb region, we generated PlasmArts from DNA mixtures containing different ratios of DNA from a child with a maternal 2211.2 deletion and DNA from the father. Samples containing only the father's DNA were used as negative controls. Eight samples at each concentration were analyzed.

Massively Multiplexed PCR and DNA Sequencing

Massively multiplex PCR and DNA sequencing methods below were used to determine allele counts at a plurality of polymorphic loci with 3-6 million (M) reads/sample for cell lines, 1.5-7 M reads/sample for tumour tissues, 18 M reads/ sample for FFPE-LCM samples, 6-7 M reads/sample for germline controls, and 18-25 M reads/sample for plasma. For two representative exemplary runs using the 3,168 SNP primer pair pool, an average of 20 million reads were used to obtain allele counts for plasma DNA libraries and 6 million reads were used to obtain allele counts for genomic DNA libraries from fresh-frozen human tumors. The percent of mapped reads (i.e. mapped to the human genome) on these two exemplary runs were 98% and 95%, respectively. The fraction of sequencing reads at a given locus with a particular allele (allele fraction) was the fractional abundance of the allele in a sample. These counts provided observed allele frequencies that were used by the data analysis methods provided immediately below in this Example to determine the ploidy state of a chromosome or chromosome segment of interest and/or to determine the average allelic imbalance of the sample.

Libraries were generated from the samples above. Adapters were ligated to DNA fragments and the fragments were amplified using the following protocol: 95° C., 2 min; 15×[95° C., 20 sec, 55° C., 20 sec, 68° C., 20 sec], 68° C. 2 min, 4° C. hold.

Multiplexed PCR allows simultaneous amplification of many targets in a single reaction. In this study, we targeted 3,168 SNPs, which were distributed across five chromosome arms as follows: 646 on  $1p,\,602$  on  $1q,\,541$  on  $2p,\,707$  on 2q, and 672 on the 2211.2 focal region. These genomic regions were selected for convenience from SNP panels available in our laboratory. Target SNPs had at least 10%population minor allele frequency (1000 Genomes Project data; Apr. 30, 2012 release) to ensure that a sufficient fraction would be heterozygous in any given patient. For each SNP, multiple primers were designed to have a maximum amplicon length of 75 bp and a melting temperature between 54.0-60.5° C. To minimize the likelihood of primer dimer product formation, primer interaction scores for all possible combinations of primers were calculated, and primers with high scores were eliminated. The 3,186 SNP primer pair pool all had  $\Delta G$  values greater than -4Kcal/mol. Candidate PCR assays were ranked and 3,168 assays were selected on the basis of target SNP minor-allele frequency, observed heterozygosity rate (from dbSNP), presence in HapMap, and amplicon length.

For PCR amplifications, 3,168 SNPs were amplified in a multiplex PCR reaction using one primer pair for each SNP, during 25 cycles, and sequencing barcodes were added in 12 additional cycles. Prior to sequencing, the barcoded products were pooled, purified with the QIAquick PCR Purification Kit (Qiagen), and quantified using the Qubit<sup>TM</sup> dsDNA BR Assay Kit (Life Technologies). Amplicons were sequenced

# US 11,519,035 B2

24

using an Illumina HiSeq 2500 sequencer with 1.5-7 M reads/sample for tumor tissue DNA and 18-25 M reads/sample for plasma cfDNA.

For the 3,168 SNP multiplex PCR reaction, approximately 7 ul (approx. 1200 ng) of library DNA, such as DNA 5 from a DNA library generated from plasma of a target individual, was used. The master mix included the following: 2× (twice manufacturer's recommended concentration) Qiagen master mix, 70 mM TMAC (tetramethylammonium chloride, Sigma), 2 nM each primer, and 7 ul nucleic acid 10 library (~1200 ng total library input) (20 ul total volume).

The cycling conditions for the 3,168 SNP multiplex PCR reaction were as follows: 95° C., 15 min; 25×[96° C., 30 sec; 65° C., 20 min; 72° C., 30 sec]; 72° C., 2 min; 4° C. hold.

For the barcoding reaction, a  $1\times$  master mix was prepared 15 that included the following: 1 uM forward primer (containing Illumina sequencing tag), 1 uM reverse primer (containing Illumina sequencing tag as well as internally-designed sequencing barcode), 1 ul of mmPCR product, diluted 1:2,000, and  $1\times$  Qiagen master mix. Barcoding cycling 20 conditions were as follows:  $95^{\circ}$  C., 10 min;  $12\times[95^{\circ}$  C., 30 sec;  $70^{\circ}$  C., 10 sec,  $60^{\circ}$  C., 30 sec;  $65^{\circ}$  C., 15 sec,  $72^{\circ}$  C., 15 sec,  $72^{\circ}$  C., 15 sec, 15 s

Data Analysis of Tumor Tissue Genomic DNA

For tumor tissue samples, CNVs were delineated by 25 transitions between allele frequency distributions. Regions with at least 100 SNPs that had an allele ratio statistically different from 0.50 were considered to be of interest. More specifically, the analysis focused on regions with average allele ratios of <0.45 or >0.55 for loci that are heterozygous 30 in the germline. A segmentation algorithm was used to exhaustively search DNA sequences in five chromosome arms as follows: 646 on 1p, 602 on 1q, 541 on 2p, 707 on 2q, and 672 on the 2211.2 for such regions, and iteratively selected them starting from the longest one until a region of 35 100 SNPs was reached. Once a ≥100 SNP region was determined to contain a CNV, it was further segmented by average allelic ratios with a minimum segment size of 50 SNPs if needed

Fresh frozen tissue samples from three patients with 40 breast cancer were also analyzed using Illumina CytoSNP-12 microarrays as previously described (Levy, B. et al. Genomic imbalance in products of conception: single-nucleotide polymorphism chromosomal microarray analysis. Obstetrics and gynecology 124, 202-209 (2014)).

Data Analysis of Circulating Tumor DNA

CNVs were identified by a maximum likelihood algorithm that searched for plasma CNVs in regions where the tumor sample from the same individual also had CNVs, using haplotype information deduced from the tumor 50 sample. This algorithm modeled expected allelic frequencies across a set of average allelic imbalances at 0.025% intervals for three sets of hypotheses: (1) all cells are normal (no allelic imbalance), (2) some/all cells have a homolog 1 deletion or homolog 2 amplification, or (3) some/all cells 55 have a homolog 2 deletion or homolog 1 amplification. For at least some of the analysis, modeling was performed up to 15% average allelic imbalance, although for the vast majority of samples AAI was less than or equal to 5%. The likelihood of each hypothesis was determined at each SNP 60 using a Bayesian classifier based on a beta binomial model of expected and observed allele frequencies at all heterozygous SNPs, and then the joint likelihood across multiple SNPs was calculated taking linkage of the SNP loci into consideration. The maximum likelihood hypothesis from the 65 comparison of expected to observed allele frequencies was then selected.

242

This algorithm also calculates the confidence of each CNV call by comparing the likelihoods of different hypotheses. A confidence threshold of 99.9% was used in plasma samples to minimize false positive results.

For dimorphic SNPs that have alleles arbitrarily designated 'A' and 'B', the allele ratio of the A allele is  $nA/(nA+n_B)$ , where nA and  $n_B$  are the number of sequencing reads for alleles A and B, respectively. Allelic imbalance is the difference between the allele ratios of A and B for loci that are heterozygous in the germline. This definition is analogous to that for SNVs, where the proportion of abnormal DNA is typically measured using mutant allele frequency, or  $n_m/(n_m+n_r)$ , where  $n_m$  and  $n_r$  are the number of sequencing reads for the mutant allele and the reference allele, respectively.

Allele frequency data was corrected for errors before it was used to generate individual probabilities. Errors that were corrected included allele amplification bias, ambient contamination, genotype contamination, and sequencing error. Ambient contamination refers to the contamination error across all SNPs in addition to sequencing errors, and genotype contamination refers to the additional contamination at some SNPs due to contamination from another sample. Ambient contamination and genotype contamination were determined on the same run as the on-test sample analysis by analyzing homozygous alleles in the sample. The ploidy status of a chromosomal segment was estimated using heterozygous loci for a test individual.

Best hypothesis was defined to be the one with the highest likelihood across all polymorphic loci. Likelihood at each locus was calculated using a beta binomial model of observed allele frequencies at each of the polymorphic loci, and the likelihood across a set of polymorphic loci was computed using the phase information deduced from the corresponding tumor sample.

A linear regression model was used to compare either expected AAI or tumor input DNA percentage and observed AAI determined by the CNV detection algorithm. P<0.05 was considered statistically significant. SigmaPlot 12.5 (Systat Software, San Jose, Calif.) and Matlab 7.12.0 R2011.a (MathWorks, Natick, Mass.) were used.

Accordingly, to evaluate the sensitivity and reproducibility of CoNVERGe, especially when the proportion of abnormal DNA for a CNV, or average allelic imbalance (AAI), is low, we used it to detect CNVs in DNA mixtures comprised of a previously characterized abnormal sample titrated into a matched normal sample. The mixtures consisted of artificial cfDNA, termed "PlasmArt", with fragment size distribution approximating natural cfDNA (see above). In the first pair, a son's tumor DNA sample having a 3 Mb Focal CNV deletion of the 2211.2 region was titrated into a matched normal sample from the father at between 0-1.5% total cfDNA. CoNVERGe reproducibly identified CNVs corresponding to the known abnormality with estimated AAI of >0.35% in mixtures of >0.5%+/-0.2% AAI, failed to detect the CNV in 6/8 replicates at 0.25% abnormal DNA, and reported a value of <0.05% for all eight negative control samples. The AAI values estimated by CoNVERGe showed high linearity (R2=0.940) and reproducibility (error variance=0.087). The assay was sensitive to different levels of amplification within the same sample. Based on these data a conservative detection threshold of 0.45% AAI could be used for subsequent analyses.

Two additional PlasmArt titrations, prepared from pairs of matched tumor and normal cell line samples and having CNVs on chromosome 1 or chromosome 2, were also evaluated. Among negative controls, all values were <0.45%, and high linearity (R2=0.952 for HCC1954 1p,

# US 11,519,035 B2

243

R2=0.993 for HCC1954 1q, R2=0.977 for HCC2218 2p, R2=0.967 for HCC2218 2q) and reproducibility (error variance=0.190 for HCC1954 1p, 0.029 for HCC1954 1q, 0.250 for HCC2218 2p, and 0.350 for HCC2218 2q) were observed between the known input DNA amount and that calculated by CoNVERGe. The difference in the slopes of the regressions for regions 1p and 1q of one sample pair correlates with the relative difference in copy number observed in the B-allelic frequencies (BAFs) of regions p and 1q of the same sample, demonstrating the relative 10 precision of the AAI estimate calculated by CoNVERGe.

CoNVERGe has application to a variety of sample sources including FFPE, Fresh Frozen, Single Cell, Germline control and cfDNA. We applied CoNVERGe to six human breast cancer cell lines and matched normal cell lines 15 to assess whether it can detect somatic CNVs. Arm-level and focal CNVs were present in all six tumour cell lines, but were absent from their matched normal cell lines, with the exception of chromosome 2 in HCC1143 in which the normal cell line exhibits a deviation from the 1:1 homolog 20 ratio. To validate these results on a different platform, we performed CytoSNP-12 microarray analyses, which produced consistent results for all samples. Moreover, the maximum homolog ratios for CNVs identified by CoNVERGe and CytoSNP-12 microarrays exhibited a strong 25 linear correlation (R2=0.987, P<0.001).

We next applied CoNVERGe to fresh-frozen (FF) and formalin-fixed, paraffin-embedded (FFPE) breast tumour tissue samples. In both sample types, several arm-level and focal CNVs were present; however, no CNVs were detected 30 in DNA from matched buffy coat samples. CoNVERGe results were highly correlated with those from microarray analyses of the same samples (R2=0.909, P<0.001 for CytoSNP-12 on FF; R2=0.992, P<0.001 for OncoScan on FFPE). CoNVERGe also produces consistent results on 35 small quantities of DNA extracted from laser capture microdissection (LCM) samples, for which microarray methods are not suitable.

Detection of CNVs in Single Cells with CoNVERGe

To test the limits of the applicability of this mmPCR 40 approach, we isolated single cells from the six aforementioned cancer cell lines and from a B-lymphocyte cell line that had no CNVs in the target regions. The CNV profiles from these single-cell experiments were consistent between three replicates and with those from genomic DNA (gDNA) 45 extracted from a bulk sample of about 20,000 cells. On the basis of the number of SNPs with no sequencing reads, the average assay drop-out rate for bulk samples was 0.48% (range: 0.41-0.60%), which is attributable to either synthesis or assay design failure. For single cells, the additional 50 average assay drop-out rate observed was 0.39% (range: 0.19-0.67%). For single cell assays that did not fail (i.e. no assay drop-out occurred), the average single ADO rate calculated using heterozygous SNPs only was 0.05% (range: 0.00-0.43%). Additionally, the percentage of SNPs with high 55 confidence genotypes (i.e. SNP genotypes determined with at least 98% confidence) was similar for both single cell and bulk samples and the genotype in the single cell samples matched those in the bulk sample (average 99.52%, range: 92.63-100.00%).

In single cells, allele frequencies are expected to directly reflect chromosome copy numbers, unlike in tumour samples where this may be confounded by TH and non-tumour cell contamination. BAFs of 1/n and (n-1)/n indicate n chromosome copies in a region. Chromosome copy numbers are indicated on the allele frequency plots for both single cells and matched gDNA samples.

244

Application of CoNVERGe to Plasma Samples

To investigate the ability of CoNVERGe to detect CNVs in real plasma samples, we applied our approach to cfDNA paired with a matched tumour biopsy from each of two stage II breast cancer patients and five late-stage ovarian cancer. In all seven patients, CNVs were detected in both FF tumour tissues and in the corresponding plasma samples. A total of 32 CNVs, at a level of >0.45% AAI, were detected in the seven plasma samples (range: 0.48-12.99% AAI) over the five regions assayed, which represent about 20% of the genome. Note that the presence of CNVs in plasma cannot be confirmed due to the lack of alternative orthogonal methods.

Although AAI estimates may appear correlated with BAFs in tumour, direct proportionality should not necessarily be expected due to tumour heterogeneity. For example, in sample BC5, regions that have BAFs compatible with N=11 were detecting; combining this with the AAI calculation from the plasma sample leads to estimates for c of 2.33% and 2.67% for the two regions. Estimating c using the other regions in the sample give values between 4.46% and 9.53%, which clearly demonstrates the presence of tumor heterogeneity.

Further CNV analyses of tumor tissue samples

We applied our mmPCR-NGS method described herein to plasma samples from four stage II breast cancer patients (BC1-BC4), and analyzed the concordance between CNVs detected in plasma and those detected in multiple tumor samples from each corresponding patient. Thus, we analyzed 4-6 tissue subsections from a tumor from each of four patients with breast cancer using mmPCR-NGS. All subsections for each patient had a CNV detected in at least one of the five targeted genomic regions (1p, lq, 2p, 2q, and 2211.2). A CNV was identified in at least one tumor subsection in 18/20 (90%) genomic regions. Among these 18 CNV-positive regions, 11 (61%) had a CNV detected in that particular region in all subsections.

Interestingly, different patterns of AAIs across these five chromosomal regions were observed among different tumor subsections. In patient BC1, for instance, a similar pattern of CNVs was observed for regions 2p, 2q, and 2211.2 in all four subsections, suggesting that these CNVs are clonal mutations. In contrast, only two of the four subsections had CNVs observed in the 1p region, and three of the four subsections had CNVs observed in the 1q region, suggesting that those CNVs were subclonal mutations. Similar patterns of possible clonal and subclonal CNVs were observed in patients BC3 and BC4, whereas BC2 appeared to be more homogenous.

In addition, even when a CNV was detected in all subsections for a particular patient, such as in the 1q region for patient BC3, the AAI often varied between subsections. Overall AAI patterns also differed between patients. Taken together, these findings suggest that mmPCR-NGS can be used to elucidate both intra- and inter-tumor clonal heterogeneity.

Concordance of CNVs in Tumors and Plasma cfDNA

To quantify the amount of overlap between CNVs detected in plasma cfDNA and those detected in tumor tissue gDNA, we used mmPCR-NGS to interrogate CNVs in tumor tissue samples and matching plasma samples from patients BC1-BC4. Seven of the 18 (39%) CNV-positive genomic regions identified in tumor subsections were also detected in the plasma (0.77%-5.80% AAI). Considering only the 11 clonal CNVs-those that were detected in all

# US 11,519,035 B2

245

tumor subsections-a CNV was detected in four (36%) of the patient-matched plasma samples (estimated AAI: 0.77%-5.80%).

Among the seven subclonal CNVs-those that were not observed in all subsections-we detected a CNV in 3/7 (43%) 5 of the regions (estimated AA: 1.24%-3.36%) in the corresponding cfDNA. Of note, in these three regions (BC1, chromosome 1p; BC1, chromosome 1q; and BC4, chromosome 2p), a CNV was detected in 10/14 (71%) of the matched tumor subsections. In contrast, in the other four 10 genomic regions that did not have a CNV detected in the corresponding plasma samples (BC3, chromosomal regions p, 2p, 2q, and 2211.2), we only detected a CNV in 7/24 (29%) of the tissue subsections. These data suggest that the more prevalent a subclonal CNV is within a tumor, the more 15 likely it is to be observed in cfDNA.

In the 150 genomic regions assayed in 30 negative controls, there were no CNVs with AAIs >0.45% and confidence >99.9%, which suggests that mmPCR-NGS has a low false-positive rate.

These data demonstrate that CNVs can be detected in plasma in a substantial fraction of samples, and suggest that the more prevalent a CNV is within a tumour, the more likely it is to be observed in cfDNA. Furthermore, CoNVERGe detected CNVs from a liquid biopsy that may have 25 otherwise gone unobserved in a traditional tumour biopsy.

### Example 28

This example provides details regarding certain exemplary sample preparation methods used for analysis of different types of samples. The sample preparation methods disclosed in this example, were used in other Examples provided herein, to generate nucleic acid templates spanning a plurality of SNP sites for next generation sequencing 35 reactions. From these NGS reactions, allele counts were generated at a plurality of polymorphic loci. These counts were then used by the analytical methods provided herein, to determine the ploidy state of a chromosome or chromosome segment of interest and/or to determine the average allelic 40 imbalance of a sample.

Single Cell CNV Protocol for 28,000-Plex PCR

Multiplexed PCR allows simultaneous amplification of many targets in a single reaction. Target SNPs were identified in each genomic region with 10% minimum population 45 minor allele frequency (1000 Genomes Project data; Apr. 30, 2012 release). For each SNP, multiple primers, semi-nested, were designed to have an amplicon length of a maximum length of 75 bp and a melting temperature between 54-60.5° C. Primer interaction scores for all possible combinations of 50 primers were calculated; primers with high scores were eliminated to reduce the likelihood of primer dimer product formation. Candidate PCR assays were ranked and selected on the basis of target SNP minor allele frequency, observed heterozygosity rate (from dbSNP), presence in HapMap, and 555 amplicon length.

In certain experiments, single cell samples were prepared and amplified using a mmPCR 28,000-plex protocol. The samples were prepared in the following way: For analysis of a single cell, cells were serial diluted until there were 3 or 60 4 cells per droplet. An individual cell was pipetted and placed into a PCR tube. The cell was lysed using Protease K, salt, and DTT using the following conditions: 56° C. for 20 minutes, 95° C. for 10 minutes, and then a 4° C. hold. For analysis of genomic DNA, DNA from the same cell line as 65 the analyzed single cell was either purchased or obtained by growing the cells and extracting the DNA. The DNA was

246

amplified in a 40 uL reaction volume containing Qiagen mp-PCR master mix (2×MM final conc), 7.5 nM primer conc. for 28K primer pairs having hemi-nested Rev primers under the following conditions: 95 C 10 min, 25×[96 C 30 sec, 65 C 29 min, 72 C 30 sec], 72 C 2 min, 4 C hold. The amplification product was diluted 1:200 in water and 2 ul added to STAR 2 (10 ul reaction volume) 1×MM, 5 nM primer conc. and PCR was performed using hemi-nested inner Fwd primer and tag specific Rev primer: 95 C 15 min, 25×[94 C 30 sec, 65 C 1 min, 60 C 5 min, 65 C 5 min, 72 C 30 sec], 72 C 2 min, 4 C hold.

Full sequence tags and barcodes were attached to the amplification products and amplified for 9 cycles using adaptor specific primers. Prior to sequencing, the barcoded library product were pooled, purified with the QIAquick PCR Purification Kit (Qiagen), and quantified using the Qubit□dsDNA BR Assay Kit (Life Technologies). Amplicons were sequenced using an Illumina HiSeq 2500 sequencer.

20 Extraction of DNA from a Blood/Plasma Sample

Blood samples were collected into EDTA tubes. The whole blood sample was centrifuged and separated into three layers: the upper layer, 55% of the blood sample, was plasma and contains cell-free DNA (cfDNA); the buffy coat middle layer contained leucocytes having DNA, <1% of total; and the bottom layer, 45% of the collected blood sample, contained erythrocytes, no DNA was present in this fraction as erythrocytes are enucleated. Circulating tumor DNA was isolated from at least 1 mL plasma using the QIAamp Circulating Nucleic Acid Kit, Qia-Amp (Qiagen, Valencia, Calif.), according to the manufacture's protocol. In certain experiments genomic DNA (gDNA) from FF tumor tissues, blood, and buccal samples was extracted using the DNeasy Blood and Tissue Kit (Qiagen).

Plasma CNV Protocol for 3,168-Plex for Chromosomes 1p, 1q, 2p, 2q, and 22q11

Plasma DNA libraries were prepared and amplified using a mmPCR 3,168-plex protocol. The samples were prepared in the following way: Up to 20 mL of blood was centrifuged to isolate the buffy coat and the plasma. Plasma extraction of cfDNA and library preparation was performed. DNA was eluted in 50 uL TE buffer. The input for mmPCR was 6.7 uL of amplified and purified Natera plasma library at an input amount of approximately 1200 ng. The plasma DNA was amplified in a 20 uL reaction volume containing Qiagen mp-PCR master mix (2×MM final conc), 2 nM tagged primer conc. (total 12.7 uM) in 3,168-plex primer pools and PCR amplified: 95 C 10 min, 25×[96 C 30 sec, 65 C 20 min, 72 C 30 sec], 72 C 2 min, 4 C hold. The amplification product was diluted 1:2,000 in water and 1 ul added to the Barcoding-PCR in a 10 uL reaction volume. The barcodes were attached to the amplification products via PCR amplification for 12 cycles using tag specific primers. Products of multiple samples were pooled and then purified with QIAquick PCR Purification Kit (Qiagen) and eluted in 50 ul DNA suspension buffer. Samples were sequenced by NGS as described for the Single Cell CNV Protocol for 28,000-plex

Breast Cancer Feasibility SNV Panel from Plasma

cfDNA from breast cancer patient blood samples was prepared and amplified using 336 primer pairs that were distributed into four 84-plex pools. Natera plasma libraries were prepared as described for Plasma CNV Protocol for 3,168-plex for Chromosomes 1p, lq, 2p, 2q, and 22q11. DNA was eluted in 50 uL TE buffer. The input for mPCR was 2.5 uL of amplified and purified Natera plasma library at an input amount of approximately 600 ng. SNPs were

# US 11,519,035 B2

### 247

selected from the 1000 Genomes map for Humans, Group 19 and dbSNP to pick targets, but only SNPs from the 1000 Genomes were used to screen for minor allele frequencies. The plasma DNA was amplified in four parallel reactions of 84-plex primer pools, a 10 uL reaction volume containing Qiagen mp-PCR master mix (2×MM final conc.), 4 mM EDTA, 7.5 nM primer concentration (total 1.26 uM) and PCR amplified: 95 C 15 min, 25×[94 C 30 sec, 65 C 15 min, 72 C 30 sec], 72 C 2 min, 4 C hold. The amplification product of the 4 subpools were each diluted 1:200 in water and 1 ul added to the Barcoding-PCR reaction in a 10 uL reaction volume containing Q5 HS HF master mix (1× final), and 1 uM each barcoding primer and each of the pools were amplified in the following reaction: 98 C 1 min, 25×[98 C 10 sec, 70 C 10 sec, 60 C 30 sec, 65 C 15 sec, 72 C 15 sec], 72 C 2 min, 4 C hold. Libraries were purified with QIAquick PCR Purification Kit (Qiagen) and eluted in 50 ul DNA suspension buffer. Samples were sequenced by paired end sequencing.

### Example 29

This example demonstrates that by using low primer concentrations such that primer amount is the limiting 25 reactant in multiplex PCR in a workflow that is followed by next generation sequencing, uniformity of density of reads, and therefore limits of detection, across a pool of amplification reactions is improved. Some experiments were carried out for plasma CNV using the 3,168-plex panel according to Example 28 above except that the total reaction volume was 10 uL instead of 20 uL. Furthermore, PCR was carried out for 15, 20, or 25 cycles. Other experiments were carried out using the four 84-plex pools on breast cancer samples according to the protocol of Example 28 except that primer concentrations were 2 nM and PCR amplification was carried out for 15, 20, or 25 cycles.

Not to be limited by theory, it is believed that primer limited multiplex PCR provides improved depth of read uniformity for multiplex PCR before multi-read sequencing, 40 such as high throughput or massively parallel sequencing, such as sequencing on an Illumina HiSeq or MiSeq system or an Ion Torrent PGM or Proton system, based on the following considerations: If some of the amplifications in a multiplex PCR have lower efficiencies than others, then with 45 normal multiplex PCR we will end up with a wide range of depth of read ("DOR") values. However, if the amount of primer is limited, and the multiplex PCR is cycled more times than what it takes to exhaust the primers, then the more efficient amplifications will stop doubling (because they 50 have no more primers to use) and the less efficient ones will continue to double; this will result in a more similar amount of amplification product for all of the amplification products. This will translate into a much more uniform distribution of

The following calculations are used to determine the number of cycles that would exact a given amount of primer and starting nucleic acid template:

assume a given starting DNA input level: 100 k copies of each target (10^5; this is easily achieved with using 60 amplified library)

assume we use 2 nM of each primer as an exemplary concentration, although other concentrations such as, for example, 0.2, 0.5, 1, 1.5, 2, 2.5, 5, or 10 nM could work too.

calculate the number of primer molecules for each primer:  $2*_{10.4}\text{-9}$  (molar concentration, 2 nM)×10\*10\*-6 (reac-

### 248

tion volume, 10 ul)×6\*10^423 (number of molecules per mole, Avogadro's number)=12\* $_{10\mathcal{A}}9$ 

calculate the amplification fold needed to consume all primers: 12\*10^9 (number of primer molecules)/10^5 (number of copies of each target)=12\*10^4

calculate the number of cycles needed to achieve this amplification fold, assuming 100% efficiency at each cycle: log 2(12\*10^4)=17 cycles. (this is log 2 because at each cycle, the number of copies doubles).

So for these conditions (100 k copies input, 2 nM primers, 10 ul reaction volume, assuming 100% PCR efficiency at each cycle), the primers would be consumed after 17 PCR cycles.

However, the key assumption is that some of the products DO NOT have 100% efficiency, so without measuring their efficiencies (which is only practicable for a small number of them anyway), it would take more than 17 cycles to consume them.

For each of four 84-plex SNV PCR primer pools we observed improved DOR efficiency with increasing cycles from 15 to 20 to 25. Similar results were obtained for experiments using the 3,168-plex panel. The limit of detection decreased (i.e. SNV sensitivity increased) with increasing depth of read. Furthermore, the sensitivity was consistently better when detecting transversion mutations than transition mutations. It is likely that additional increases in DOR efficiency can be obtained with additional cycles when using primer-limiting multiplex PCR before multi-read sequencing.

Accordingly, in one aspect provided herein is a method of amplifying a plurality of target loci in a nucleic acid sample that includes (i) contacting the nucleic acid sample with a library of primers and other primer extension reaction components to provide a reaction mixture, wherein the relative amount of each primer in the reaction mixture compared to the other primer extension reaction components creates a reaction wherein the primers are present at a limiting concentration, and wherein the primers hybridize to a plurality of different target loci; and (ii) subjecting the reaction mixture to primer extension reaction conditions for sufficient number of cycles to consume or exhaust the primers in the library of primers, to produce amplified products that include target amplicons. For example, the plurality of different target loci can include at least 2, 3, 5, 10, 25, 50, 100, 200, 250, 500, 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci, and at most, 50, 100, 200, 250, 500, 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; 100,000, 200,000, 250,000, 500, 000, and 1,000,000 different target loci to produce a reaction mixture.

The method in illustrative embodiments, includes determining an amount of primer that will be a rate limiting amount. This calculation typically includes estimating and/ or determining the number of target molecules and involves analyzing and/or determining the number of amplification cycles performed. For example, in illustrative embodiments, the concentration of each primer is less than 100, 75, 50, 25, 10, 5, 2, 1, 0.5, 0.25, 0.2 or 0.1 nM. In various embodiments, the GC content of the primers is between 30 to 80%, such as between 40 to 70% or 50 to 60%, inclusive. In some embodiments, the range of GC content (e.g., the maximum GC content minus minimum GC content, such as 80%-60%=a range of 20%) of the primers is less than 30, 20, 10, 65 or 5%. In some embodiments, the melting temperature  $(T_m)$ of the primers is between 40 to 80° C., such as 50 to 70° C. 55 to 65° C., or 57 to 60.5° C., inclusive. In some embodi-

# US 11,519,035 B2

249

ments, the range of melting temperatures of the primers is less than 20, 15, 10, 5, 3, or 1° C. In some embodiments, the length of the primers is between 15 to 100 nucleotides, such as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, 20 to 65 nucleotides, inclusive. In some embodiments, the primers include a tag that is not target specific, such as a tag that forms an internal loop structure. In some embodiments, the tag is between two DNA binding regions. In various embodiments, the primers include a 5' region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In various embodiments, the length of the 3' region is at least 7 nucleotides. In some embodiments, the length of the 3' region is between 7 and 20 nucleotides, such as between 7 to 15 nucleotides, or 7 to 10 nucleotides, inclusive. In various embodiments, the test primers include a 5' region that is not specific for a target locus (such as a tag or a universal primer binding site) followed by a region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In some embodiments, the range of the length of the primers is less than 50, 40, 30, 20, 10, or 5 nucleotides. In some embodiments, the length of the target amplicons is between 50 and 25 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 100, 75, 50, 25, 15, 10, or 5 nucleotides.

250

In various embodiments of any of the aspects of the invention, the primer extension reaction conditions are polymerase chain reaction conditions (PCR). In various embodiments, the length of the annealing step is greater than 3, 5, 8, 10, or 15 minutes but less than 240, 120, 60, or 30 minutes. In various embodiments, the length of the extension step is greater than 3, 5, 8, 10, or 15 minutes but less than 240, 120, 60 or 30 minutes.

All patents, patent applications, and published references cited herein are hereby incorporated by reference in their entirety. While the methods of the present disclosure have been described in connection with the specific embodiments thereof, it will be understood that it is capable of further modification. Furthermore, this application is intended to cover any variations, uses, or adaptations of the methods of the present disclosure, including such departures from the present disclosure as come within known or customary practice in the art to which the methods of the present disclosure pertain, and as fall within the scope of the appended claims. For example, any of the methods disclosed herein for DNA can be readily adapted for RNA by including a reverse transcription step to convert the RNA into DNA. Examples that use polymorphic loci for illustration can be readily adapted for the amplification of nonpolymorphic loci if desired. Any of the methods disclosed herein can be used with a low level of multiplexing if desired (such as with less than 1,000, 750, 500, 250, 100, 75, 50, 25, or 10 different primers or different primer pairs to different target loci).

# SEQUENCE LISTING

The patent contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (https://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US11519035B2). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

What is claimed is:

1. A method for amplifying and sequencing DNA, comprising:

tagging isolated cell free DNA with one or more universal tail adaptors to generate tagged products, wherein the isolated cell-free DNA is isolated from a blood sample collected from a subject who is not a pregnant women; amplifying the tagged products one or more times to generate final amplification products, wherein one of the amplification steps comprises targeted amplification of a plurality of single nucleotide polymorphism (SNP) loci in a single reaction volume, wherein one of the amplifying steps introduces a barcode and one or more sequencing tags; and

sequencing the plurality of SNP loci on the cell free DNA by conducting massively parallel sequencing on the final amplification products, wherein the plurality of 60 SNP loci comprises 25-2,000 loci associated with cancer.

- 2. The method of claim 1, wherein tagging the cell free DNA comprises ligating the one or more universal tail adaptors to the cell free DNA.
- 3. The method of claim 2, wherein the one or more universal tail adaptors each comprise a first strand and a

- second strand, wherein a first end of each of the universal tail adaptors comprises a double-stranded section comprising the 5' portion of the first strand and the 3' portion of the second strand, wherein the first end is ligated to the cell free DNA.
- 4. The method of claim 3, wherein amplifying the tagged products comprises a first amplifying step and a second amplifying step, wherein the first amplifying step comprises using a first target-specific primer that specifically anneals to a target sequence and a first adaptor primer having a nucleotide sequence identical to a first portion of the first strand to generate a first amplification product.
- 5. The method of claim 4, wherein the second amplifying step comprises using a second target-specific primer that specifically anneals to the first amplification product and a second adaptor primer having a nucleotide sequence identical to a second portion of the first strand to generate the final amplification product.
- 6. The method of claim 5, wherein the second adaptor primer is nested relative to the first adaptor primer.
- 7. The method of claim 6, wherein the second target-specific primer comprises an index tag.
- **8.** The method of claim **7**, wherein the second amplifying step further comprises using an index primer comprising a sequence complementary to the index tag.

Case: 24-1324 Document: 42-1 Page: 248 Filed: 03/18/2024

US 11,519,035 B2

252

- 9. The method of claim 8, wherein the index primer comprises the barcode and a first sequencing tag.
- 10. The method of claim 1, wherein the one or more universal tail adaptors comprise a second barcode.
- 11. The method of claim 1, wherein the one or more 5 universal tail adaptors comprise a second sequencing tag.
- 12. The method of claim 1, wherein the one or more universal tail adaptors comprise a first universal tail adaptor
- and a second universal tail adaptor. 13. The method of claim 12, wherein tagging the cell free 10 DNA comprises amplifying the cell free DNA with a first primer comprising the first universal tail adaptor and a second primer comprising the second universal tail adaptor.
- 14. The method of claim 12, wherein amplifying the
- tagged products comprises a single amplifying step.

  15. The method of claim 14, wherein amplifying the tagged products comprises using a third primer and a fourth primer, wherein the third primer comprises a first sequencing tag and wherein the fourth primer comprises a second sequencing tag.

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 **PR的音: 249** Filed: 03/18/2024

Query Reports <u>U</u>tilities Help Log Out

18BD, APPEAL, MEDIATION

# U.S. District Court North Carolina Middle District (NCMD) CIVIL DOCKET FOR CASE #: 1:23-cv-00629-CCE-JLW

NATERA, INC. v. NEOGENOMICS LABORATORIES, INC.

Assigned to: CHIEF JUDGE CATHERINE C. EAGLES

Referred to: MAG/JUDGE JOE L. WEBSTER

Cause: 35:271 Patent Infringement

Date Filed: 07/28/2023 Jury Demand: Both Nature of Suit: 830 Patent Jurisdiction: Federal Question

# **Plaintiff**

NATERA, INC.

# represented by KEVIN P.B. JOHNSON

QUINN EMANUEL URQUHART & SULLIVAN, LLP

555 TWIN DOLPHIN DRIVE

5TH FLOOR

REDWOOD SHORES, CA 94065

650-801-5000 Fax: 650-801-5100

Email: kevinjohnson@quinnemanuel.com

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

# VICTORIA MAROULIS

QUINN EMANUEL 555 TWIN DOLPHIN DR. STE 5TH FLOOR REDWOOD SHORES, CA 94065

650-801-5000

Email:

victoriamaroulis@quinnemanuel.com LEAD ATTORNEY ATTORNEY TO BE NOTICED

# ANDREW J BRAMHALL

QUINN EMANUEL URQUHART & SULLIVAN, LLP 555 TWIN DOLPHIN DRIVE 5TH FLOOR REDWOOD SHORES, CA 94065

650-801-5000 Fax: 650-801-5100

Email:

andrewbramhall@quinnemanuel.com ATTORNEY TO BE NOTICED

# ANDREW R. SHORES

**WILLIAMS MULLEN** 

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 PM EGGE: 250 Filed: 03/18/2024

301 FAYETTEVILLE STREET SUITE 1700 RALEIGH, NC 27601 919-981-4015 Email: ashores@williamsmullen.com ATTORNEY TO BE NOTICED

# **CARMELLE F. ALIPIO**

WILLIAMS MULLEN
301 FAYETTEVILLE STREET
SUITE 1700
RALEIGH, NC 27601
919-981-4038
Fax: 919-981-4300
Email: calipio@williamsmullen.com
ATTORNEY TO BE NOTICED

# **JOSHUA HIRAM HARRIS**

QUINN EMANUEL LLP 191 N. WACKER DRIVE CHICAGO, IL 60606 313-330-2338 Email: joshua.harris8@icloud.com ATTORNEY TO BE NOTICED

# **KEVIN ALEXANDER SMITH**

QUINN EMANUEL URQUHART & SULLIVAN LLP 50 CALIFORNIA ST. STE 22ND FLOOR SAN FRANCISCO, CA 94111 415-876-6600

Fax: 415-875-6700 Email: kevinsmith@quinnemanuel.com

ATTORNEY TO BE NOTICED

# **SANDRA HABERNY**

QUINN EMANUEL 865 S. FIGUEROA STREET STE 10TH FLOOR LOS ANGELES, CA 90017 213-443-3438

Fax: 213-443-3100

Email: sandrahaberny@quinnemanuel.com

ATTORNEY TO BE NOTICED

# TARA SRINIVASAN

QUINN EMANUEL URQUHART & SULLIVAN LLP 50 CALIFORNIA ST. FLOOR 22 SAN FRANCISCO, CA 94111 510-378-2977

Email: tarasrinivasan@quinnemanuel.com

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 PMage: 251 Filed: 03/18/2024

ATTORNEY TO BE NOTICED

## **ROBERT VAN ARNAM**

WILLIAMS MULLEN 301 FAYETTEVILLE STREET SUITE 1700

RALEIGH, NC 27601 919-981-4055 Fax: 919-981-4300

Email: rvanarnam@williamsmullen.com

ATTORNEY TO BE NOTICED

V.

# **Defendant**

NEOGENOMICS LABORATORIES, INC.

# represented by JOHN F. MORROW, JR.

WOMBLE BOND DICKINSON (US) LLP ONE WEST FOURTH STREET WINSTON-SALEM, NC 27101 336-721-3584 Fax: 336-733-8429

Email: John.Morrow@wbd-us.com
LEAD ATTORNEY

ATTORNEY TO BE NOTICED

# **AUGUST MELCHER**

WEIL, GOTSHAL & MANGES LLP 201 REDWOOD SHORES PARKWAY STE 400 REDWOOD SHORES, CA 94065 650-802-3285 Email: august.melcher@weil.com ATTORNEY TO BE NOTICED

# DARALYN J. DURIE

MORRISON FOERSTER LLP 425 MARKET STREET SAN FRANCISCO, CA 94105 415-268-7000 Fax: 415-268-7522 Email: ddurie@mofo.com

ATTORNEY TO BE NOTICED

# **DEREK WALTER**

WEIL, GOTSHAL & MANGES LLP 201 REDWOOD SHORES PARKWAY REDWOOD SHORES, CA 94065 650-802-3934 Email: derek.walter@weil.com ATTORNEY TO BE NOTICED

# EDWARD R. REINES

WEIL, GOTSHAL & MANGES LLP

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 PM EGGE: 252 Filed: 03/18/2024

201 REDWOOD SHORES PARKWAY REDWOOD SHORES, CA 94065

650-802-3022 Fax: 650-802-3100

Email: edward.reines@weil.com ATTORNEY TO BE NOTICED

# **ELIZABETH RYAN**

WEIL, GOTSHAL & MANGES LLP 200 CRESCENT COURT SUITE 300 DALLAS, TX 75201 214-746-8158 Fax: 214-746-7777 Email: liz.ryan@weil.com

ATTORNEY TO BE NOTICED

# JOHN J NOLAN

WEIL GOTSHAL & MANGES LLP 767 FIFTH AVENUE NEW YORK, NY 10153 212-310-8000 Email: jack.nolan@weil.com ATTORNEY TO BE NOTICED

# JOHN D. WOOTEN, IV

WOMBLE BOND DICKINSON (US) LLP 300 N. GREENE STREET SUITE 1900 GREENSBORO, NC 27401 336-574-8090 Fax: 336-574-4521 Email: JD.Wooten@wbd-us.com ATTORNEY TO BE NOTICED

# **MATTHEW I KREEGER**

MORRISON & FOERSTER 425 MARKET STREET SAN FRANCISCO, CA 94105 415-268-6467 Fax: 415-268-7522 Email: mkreeger@mofo.com ATTORNEY TO BE NOTICED

# PATRICK ROBERT LYONS

WEIL GOTSHAL & MANGES LLP 767 FIFTH AVENUE NEW YORK, NY 10153 212-310-8486 Email: patrick.lyons@weil.com ATTORNEY TO BE NOTICED

# **Counter Claimant**

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 PMage: 253 Filed: 03/18/2024

NEOGENOMICS LABORATORIES, INC.

represented by **JOHN F. MORROW , JR.**(See above for address) *LEAD ATTORNEY* 

ATTORNEY TO BE NOTICED

#### **AUGUST MELCHER**

(See above for address)

ATTORNEY TO BE NOTICED

## **DEREK WALTER**

(See above for address)

ATTORNEY TO BE NOTICED

#### **EDWARD R. REINES**

(See above for address)

ATTORNEY TO BE NOTICED

# **ELIZABETH RYAN**

(See above for address)

ATTORNEY TO BE NOTICED

#### JOHN J NOLAN

(See above for address)

ATTORNEY TO BE NOTICED

## PATRICK ROBERT LYONS

(See above for address)

ATTORNEY TO BE NOTICED

V.

**Counter Defendant** 

NATERA, INC.

# represented by KEVIN P.B. JOHNSON

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

# **VICTORIA MAROULIS**

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

## ANDREW J BRAMHALL

(See above for address)

ATTORNEY TO BE NOTICED

# ANDREW R. SHORES

(See above for address)

ATTORNEY TO BE NOTICED

# **CARMELLE F. ALIPIO**

(See above for address)

ATTORNEY TO BE NOTICED

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 PMEGE: 254 Filed: 03/18/2024

#### **KEVIN ALEXANDER SMITH**

(See above for address)

ATTORNEY TO BE NOTICED

#### **SANDRA HABERNY**

(See above for address)

ATTORNEY TO BE NOTICED

# **ROBERT VAN ARNAM**

(See above for address)

ATTORNEY TO BE NOTICED

## **Counter Claimant**

NEOGENOMICS LABORATORIES, INC.

# represented by JOHN F. MORROW, JR.

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

#### **AUGUST MELCHER**

(See above for address)

ATTORNEY TO BE NOTICED

#### DEREK WALTER

(See above for address)
ATTORNEY TO BE NOTICED

#### EDWARD R. REINES

(See above for address)

ATTORNEY TO BE NOTICED

## **ELIZABETH RYAN**

(See above for address)

ATTORNEY TO BE NOTICED

# **JOHN J NOLAN**

(See above for address)

ATTORNEY TO BE NOTICED

# PATRICK ROBERT LYONS

(See above for address)

ATTORNEY TO BE NOTICED

V.

# **Counter Defendant**

NATERA, INC.

# represented by KEVIN P.B. JOHNSON

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

# VICTORIA MAROULIS

(See above for address)

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 Prespe: 255 Filed: 03/18/2024

LEAD ATTORNEY ATTORNEY TO BE NOTICED

## ANDREW J BRAMHALL

(See above for address)

ATTORNEY TO BE NOTICED

#### ANDREW R. SHORES

(See above for address)

ATTORNEY TO BE NOTICED

#### **CARMELLE F. ALIPIO**

(See above for address)

ATTORNEY TO BE NOTICED

# KEVIN ALEXANDER SMITH

(See above for address)

ATTORNEY TO BE NOTICED

#### SANDRA HABERNY

(See above for address)

ATTORNEY TO BE NOTICED

## TARA SRINIVASAN

(See above for address)

ATTORNEY TO BE NOTICED

#### **ROBERT VAN ARNAM**

(See above for address)

ATTORNEY TO BE NOTICED

# **Counter Claimant**

NEOGENOMICS LABORATORIES, INC.

# represented by JOHN F. MORROW, JR.

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

## **AUGUST MELCHER**

(See above for address)

ATTORNEY TO BE NOTICED

#### **DEREK WALTER**

(See above for address)
ATTORNEY TO BE NOTICED

#### **EDWARD R. REINES**

(See above for address)

ATTORNEY TO BE NOTICED

# **ELIZABETH RYAN**

(See above for address)

ATTORNEY TO BE NOTICED

#### **JOHN J NOLAN**

2/7/24, 12:10 PM Case: 24-1324 Document: 42-1 PM 256 Filed: 03/18/2024

(See above for address)

ATTORNEY TO BE NOTICED

# PATRICK ROBERT LYONS

(See above for address)
ATTORNEY TO BE NOTICED

V.

## **Counter Defendant**

NATERA, INC.

# represented by KEVIN P.B. JOHNSON

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

## VICTORIA MAROULIS

(See above for address)

LEAD ATTORNEY

ATTORNEY TO BE NOTICED

# ANDREW R. SHORES

(See above for address)
ATTORNEY TO BE NOTICED

#### **CARMELLE F. ALIPIO**

(See above for address)

ATTORNEY TO BE NOTICED

# **JOSHUA HIRAM HARRIS**

(See above for address)
ATTORNEY TO BE NOTICED

## **KEVIN ALEXANDER SMITH**

(See above for address)

ATTORNEY TO BE NOTICED

# SANDRA HABERNY

(See above for address)

ATTORNEY TO BE NOTICED

# TARA SRINIVASAN

(See above for address)

ATTORNEY TO BE NOTICED

## **ROBERT VAN ARNAM**

(See above for address)
ATTORNEY TO BE NOTICED

<b>Date Filed</b>	#	Docket Text
07/28/2023		COMPLAINT against Neogenomics Laboratories, Inc. (Filing fee \$ 402 receipt number ANCMDC-3618904.), filed by Natera, Inc. (Attachments: # 1 Exhibit 1. US 11530454, # 2 Exhibit 2. US 11519035, # 3 Exhibit 3. 454 Patent Claim Chart, # 4 Exhibit 4. 035

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1
		Patent Claim Chart, # 5 Exhibit 5. RaDaR Residual Disease and Recurrence, # 6 Exhibit 6. Heider AACR Virtual Poster, # 7 Exhibit 7. AACR 2020, # 8 Exhibit 8. Lipsyc Sharf, # 9 Exhibit 9. Gale, # 10 Exhibit 10. Flach, # 11 Exhibit 11. AACR 2020 Poster 3097, # 12 Exhibit 12. Cutts AACR 2021, # 13 Exhibit 13. Gale Pub, # 14 Exhibit 14. AACR 2021 Pubs, # 15 Exhibit 15. TAm-Seq Poster, # 16 Exhibit 16. Forshew, # 17 Exhibit 17. NeoGenomics Labs, # 18 Exhibit 18. CAP Certificate, # 19 Exhibit 19. CLIA Certificate, # 20 Exhibit 20. Sebastian, # 21 Exhibit 21. NeoGenomics Annual Report, # 22 Exhibit 22. RaDaR Discussion Guide, # 23 Exhibit 23. RaDaR Commercial Avail Announcement, # 24 Exhibit 24. MoIDX-DEX Diagnostics, # 25 Exhibit 25. RaDaR Assay Request Form, # 26 Exhibit 26. NeoGenomics Global Conf. Takeaways, # 27 Exhibit 27. Top Ten Article, # 28 Exhibit 28. Signatera - Innovation Award, # 29 Exhibit 29. MedTech Breakthrough winners, # 30 Exhibit 30. RaDaR Medicare Coverage Article, # 31 Civil Cover Sheet)(VAN ARNAM, ROBERT) (Entered: 07/28/2023)
07/28/2023	<u>2</u>	NOTICE of Attorney Appearance by attorney ROBERT VAN ARNAM on behalf of Plaintiff Natera, Inc (VAN ARNAM, ROBERT) (Entered: 07/28/2023)
07/28/2023	3	NOTICE of Attorney Appearance by attorney ANDREW R. SHORES on behalf of Plaintiff Natera, Inc (SHORES, ANDREW) (Entered: 07/28/2023)
07/28/2023	4	Corporate Disclosure Statement. (SHORES, ANDREW) (Main Document 4 replaced on 8/1/2023 with flattened image.) (Hicks, Samantha) (Entered: 07/28/2023)
07/31/2023	<u>5</u>	MOTION for Preliminary Injunction by Natera, Inc. Response to Motion due by 8/21/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	6	MEMORANDUM filed by Plaintiff Natera, Inc re 5 MOTION for Preliminary Injunction filed by Natera, Inc. (Attachments: # 1 Exhibit 1 - NeoGenomics Gets Medicare Coverage, # 2 Exhibit 2 - Motion to Dismiss for Failure to State a Claim, # 3 Exhibit 3 - Opposition Brief to Inivata MTD, # 4 Exhibit 4 - Reply Brief re MTD, # 5 Exhibit 5 - 2022 NeoGenomics Laboratories Annual Report, # 6 Exhibit 6 - Strategic Collaboration and Investment from NeoGenomics, # 7 Exhibit 7 - MolDX - DEX Diagnostics Exchange, # 8 Exhibit 8 - RaDaR Assay Request Form, # 9 Exhibit 9 - Medicare Extends Coverage of Natera Signatera MRD Test, # 10 Exhibit 10 - Neo Q1 2023 Earnings call, # 11 Exhibit 11 - BTIG - Natera, # 12 Exhibit 12 - SVB Securities, # 13 Exhibit 13 - MRD The Unlocking of Billions in TAM, # 14 Exhibit 14 - Seeking Alpha - Natera, Inc., # 15 Exhibit 15 - Significant Ramp Ahead in Cancer Liquid Biopsy, # 16 Exhibit 16 - Refinitiv StreetEvents)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	7	DECLARATION of Anup Malani, Ph.D. filed by Plaintiff Natera, Inc re 5 MOTION for Preliminary Injunction filed by Natera, Inc. (Attachments: #1 Exhibit 1 - 35 USC 283, #2 Exhibit 2 - Phylogenetic ctDNA analysis, #3 Exhibit 3 - Abbott Laboratories v. Sandoz, Inc., #4 Exhibit 4 - Understanding clonoSEQ, #5 Exhibit 5 - Altana & Wyeth v Teva, #6 Exhibit 6 - Amazon.com, Inc. v. Barnesandnoble.com, #7 Exhibit 7 - Explaining Molecular Residual Disease, #8 Exhibit 8 - Apple v. Samsung, #9 Exhibit 9 - Apple v. Samsung, #10 Exhibit 10 - How a Simple Blood Test Could Reduce Cancer Deaths, #11 Exhibit 11 - Automobile Prices in Market Equilibrium, #12 Exhibit 12 - Understanding MRD in Clinical Trials, #13 Exhibit 13 - Broadcom Corp. v. Emulex Corp, #14 Exhibit 14 - Compelling data sets at AACR, #15 Exhibit 15 - Key Takeaways from our MRD, #16 Exhibit 16 - BTIG - Natera, #17 Exhibit 17 - BTIG - Natera, #18 Exhibit 18 - BTIG - Natera, #19 Exhibit 21 - OncoCyte Corporation, #20 Exhibit 20 - NEOs operational turnaround, #21 Exhibit 21 - OncoCyte Corporation, #22 Exhibit 22 - We remain positive on MRD testing, #23 Exhibit 23 - NTRA is a Pillar, #24 Exhibit 24 - Diagnostic Value of Liquid Biopsy in the Era of Precision, #25 Exhibit 25 - Celsis in Vitro, Inc. v. CellzDirect, Inc.,, #26 Exhibit 26 - Local Coverage Determination, #27 Exhibit 27 -

2/7/24, 12:10 PM	Case: 24-1324 Document: 42-1 @@e: 258 Filed: 03/18/2024
	LDT and CLIA FAQs, # 28 Exhibit 28 - Medicare Program Integrity Manual, # 29 Exhibit 29 - MolDx Minimal Residual Disease Testing, # 30 Exhibit 30 - Proprietary Gross Margin Analysis, # 31 Exhibit 31 - Liquid Biopsy Report, # 32 Exhibit 32 - Natera, Inc. Q4 Upside, # 33 Exhibit 33 - Diaceutics, # 34 Exhibit 34 - Douglas Dynamics, LLC v. Buyers Products Co., # 35 Exhibit 35 - E.I. du Pont de Nemours v. Polaroid Grap, # 36 Exhibit 36 - eBay Inc. v. MercExchange LLC, # 37 Exhibit 37 - Economic Forecasting, # 38 Exhibit 38 - Molecular Residual Disease, # 39 Exhibit 39 - FDA Website, # 40 Exhibit 40 - Distribution of IVD Products)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	NOTICE by Natera, Inc re 7 Declaration. (Attachments: # 1 Exhibit 41 - Principles for Codevelopment, # 2 Exhibit 42 - Foundation Medicine, # 3 Exhibit 43 - Preemption, Leapfrogging, and Competition in Patent Races, # 4 Exhibit 44 - Inivata Expects New Minimal Residual Disease, # 5 Exhibit 45 - NeoGenomics Gets Medicare Coverage for Radar, # 6 Exhibit 46 - Detection of Circulating Tumor DNA, # 7 Exhibit 47 - Preemtive Patenting, # 8 Exhibit 48 - Biopharma Solutions, # 9 Exhibit 49 - Tests for Patients, # 10 Exhibit 50 - Hsieh, Mishra, Gobeli (2003), # 11 Exhibit 51 - Hypbritech v. Abbott, # 12 Exhibit 52 - Hypbritech v. Abbott, # 13 Exhibit 53 - Invitate's PCM Assay, # 14 Exhibit 54 - Probabilistic Patents, # 15 Exhibit 55 - First-Mover Advantages, # 16 Exhibit 56 - Lighter Capital, # 17 Exhibit 57 - Regulatory approval pathways, # 18 Exhibit 58 - Principles of Economics)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	NOTICE by Natera, Inc re 7 Declaration. (Attachments: # 1 Exhibit 58 - Con't, # 2 Exhibit 59 - Mayo Clinic, # 3 Exhibit 60 - MedTech Breakthrough Website, # 4 Exhibit 61 - Patent 11,530,454, # 5 Exhibit 62 - Patent 11,519,035, # 6 Exhibit 63 - Natera Inc 1Q23, # 7 Exhibit 64 - Key Takeaways, # 8 Exhibit 65 - Morgan Stanley - Natera, # 9 Exhibit 66 - FDA Grants Breakthrough Device, # 10 Exhibit 67 - FDA Grants Breakthrough Device, # 11 Exhibit 68 - Foundation Medicine, # 12 Exhibit 69 - Natera Press Release, # 13 Exhibit 70 - Medicare Extends Coverage, # 14 Exhibit 71 - Natera Announces Publication, # 15 Exhibit 72 - Natera Announces Use of Signatera, # 16 Exhibit 73 - Natera Awarded Advanced Diagnostic, # 17 Exhibit 74 - Natera Press Release, # 18 Exhibit 75 - Nateras Signatera Test)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	NOTICE by Natera, Inc re 7 Declaration. (Attachments: # 1 Exhibit 76 - Natera Press Release, # 2 Exhibit 77 - Natera - About us, # 3 Exhibit 78 - Pioneers in Individualized Genetic Testing, # 4 Exhibit 79 - Natera Website, # 5 Exhibit 80 - Natera Website, # 6 Exhibit 81 - Signatera Overview, # 7 Exhibit 82 - Natera Website, # 8 Exhibit 83 - A personalized tumor informed approach, # 9 Exhibit 84 - Company Fact Sheet, # 10 Exhibit 85 - Natera - Form 10-K (2015), # 11 Exhibit 86 - Natera - Form 10-K (2017), # 12 Exhibit 87 - Natera - Form 10-K (2020), # 13 Exhibit 88 - Natera - Form 10-K (2022), # 14 Exhibit 89 - Natera - Q1 2023, # 15 Exhibit 90 - Natera - Q4 22, # 16 Exhibit 91 - JPM HC Conf Presentation, # 17 Exhibit 92 - Natera - Q3 22, # 18 Exhibit 93 - National Cancer Institute, # 19 Exhibit 94 - National Cancer Institute, # 20 Exhibit 95 - National Cancer Institute, # 21 Exhibit 96 - NCCN Website - About, # 22 Exhibit 97 - NeoGenomics Press Release, # 23 Exhibit 98 - NeoGenomics - About, # 24 Exhibit 99 - NeoGenomics Website, # 25 Exhibit 100 - NeoGenomics Website, # 26 Exhibit 101 - NeoGenomics Website, # 27 Exhibit 102 - NeoGenomics Website, # 28 Exhibit 103 - NeoGenomics Website, # 29 Exhibit 104 - NeoGenomics Website, # 30 Exhibit 105 - NeoGenomics Website, # 31 Exhibit 106 - NeoGenomics Website, # 32 Exhibit 107 - NeoGenomics Website, # 33 Supplement 108 - NeoGenomics Website, # 34 Exhibit 110 - NeoGenomics Website, # 35 Exhibit 110 - NeoGenomics Website, # 36 Exhibit 111 - NeoGenomics Website, # 37 Exhibit 112 - NeoGenomics Website, # 38 Exhibit 113 - NeoGenomics Website, # 39 Exhibit 114 - NeoGenomics Website, # 40 Exhibit 115 - 2023 Investor Day Presentation)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)

7/24, 12:10 PM	Case: 24-1324 Document: 42-1 Prege: 259 Filed: 03/18/2024
07/31/2023	NOTICE by Natera, Inc re 7 Declaration. (Attachments: # 1 Exhibit 116 - NeoGenomics - Form 8-K, # 2 Exhibit 117 - Neogenomics - Form 10-K, # 3 Exhibit 118 - Neogenomics - Form 10-K, # 4 Exhibit 119 - Neogenomics - Form 10-K, # 5 Exhibit 120 - Neogenomics - Form 10-K, # 6 Exhibit 121 - Neogenomics - Form 10-K, # 7 Exhibit 122 - Obuchowski, Zhou, # 8 Exhibit 123 - Orthopedic Design, # 9 Exhibit 124 - MRD The Unlocking of Billions in TAM, # 10 Exhibit 125 - Hot Off The Press, # 11 Exhibit 126 - Presidio Components v. American Technical Ceramics, # 12 Exhibit 127 - Study designs Part 1, # 13 Exhibit 128 - Refinitiv StreetEvents, # 14 Exhibit 129 - Refinitiv StreetEvents, # 15 Exhibit 130 - Refinitiv StreetEvents)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	NOTICE by Natera, Inc re 7 Declaration. (Attachments: # 1 Exhibit 131 - Refinitiv StreetEvents, # 2 Exhibit 132 - Refinitiv StreetEvents, # 3 Exhibit 133 - Refinitiv StreetEvents, # 4 Exhibit 134 - Refinitiv StreetEvents, # 5 Exhibit 135 - Refinitiv StreetEvents, # 6 Exhibit 136 - Refinitiv StreetEvents, # 7 Exhibit 137 - Refinitiv StreetEvents, # 8 Exhibit 138 - Refinitiv StreetEvents, # 9 Exhibit 139 - Refinitiv StreetEvents, # 10 Exhibit 140 - First-Mover Advantages, # 11 Exhibit 141 - Natera, Inc. (NTRA) Q1 2023, # 12 Exhibit 142 - Natera, Inc. (NTRA) Q4 2022, # 13 Exhibit 143 - State Indus., Inc. v. Mor-Flo Indus, # 14 Exhibit 144 - SVB Leerink, # 15 Exhibit 145 - SVB Leerink, # 16 Exhibit 146 - SVB Leerink, # 17 Exhibit 147 - SVB Leerink, # 18 Exhibit 148 - SVB Leerink, # 19 Exhibit 149 - SVB Securities, # 20 Exhibit 150 - SVB Securities, # 21 Exhibit 151 - SVB Securities, # 22 Exhibit 152 - SVB Securities, # 23 Exhibit 153 - Going Deeper On MRD, # 24 Exhibit 154 - Testing.com, # 25 Exhibit 155 - Thomson Reuters Streetevents, # 26 Exhibit 156 - Thomson Reuters Streetevents, # 27 Exhibit 157 - Initiate Coverage of NTRA with a Buy, # 28 Exhibit 158 - Veeco Instruments v. SGL Carbon, # 29 Exhibit 159 - Wall Street Journal, # 30 Exhibit 160 - Winter v. Natural Resources Defense Council, Inc., # 31 Exhibit 161 - Introductory Econometrics)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	DECLARATION filed by Plaintiff Natera, Inc re 5 MOTION for Preliminary Injunction DR. MICHAEL L. METZKER (Redacted) filed by Natera, Inc. (Attachments: # 1 Exhibit 1 - CV, # 2 Exhibit 2 - Kirkizlar et al, # 3 Exhibit 3 - Leon1997Paper, # 4 Exhibit 4 - RaDaR Technology, # 5 Exhibit 5 - Gale, # 6 Exhibit 6 - Analytical-performance, # 7 Exhibit 7 - Forshew, # 8 Exhibit 8 - Heider AACR, # 9 Exhibit 9 - Lipsyc Sharf 2022, # 10 Exhibit 10 - Flach + Supp Methods, # 11 Exhibit 11 - Marsico AACR_2020, # 12 Exhibit 12 - 536-Cutts-AACR-2021, # 13 Exhibit 13 - Plagnol2018Paper, # 14 Exhibit 14 - AACR_2023_Evaluation, # 15 Exhibit 15 - RaDaR_PatientDiscussionGuide, # 16 Exhibit 16 - RaDaR Assay Request Form, # 17 Exhibit 17 - Oncology-Clinical, # 18 Exhibit 18 - jamaoncology, # 19 Exhibit 19 - Coombes2019Paper, # 20 Exhibit 20 - Abbosh2017Paper, # 21 Exhibit 21 - Christensen2019Paper, # 22 Exhibit 22 - Kotani2023Paper)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	DECLARATION of SOLOMON MOSHKEVICH filed by Plaintiff Natera, Inc re 5 MOTION for Preliminary Injunction <i>Redacted</i> filed by Natera, Inc. (Attachments: # 1 Exhibit A - Natera Receives Final Medicare Coverage, # 2 Exhibit B - Natera Announcement, # 3 Exhibit C - 2022 10-K Natera, # 4 Exhibit D - NeoGenomics Announces Commercial Availability of the RaDaR, # 5 Exhibit E - Neogenomics Investor Day Presentation, # 6 Exhibit F - Email (under seal), # 7 Exhibit G - Power Point, # 8 Exhibit H - Refinitiv StreetEvents, # 9 Exhibit I - Refinitiv StreetEvents, # 10 Exhibit J - Russell - AACR 2023, # 11 Exhibit K - 2023_Validation, # 12 Exhibit L - Abbosh 2017, # 13 Exhibit M - Lebow 2022, # 14 Exhibit N - NSCLC clinical data)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	MOTION to Seal unredacted versions of the Declaration of Dr. Michael L. Metzker, the Declaration of Solomon Moshkevich, and the entirety of Exhibit F attached to the

7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 @@e: 260 Filed: 03/18/2024
		Moshkevich Declaration [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by Natera, Inc. Response to Motion due by 8/14/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	16	MEMORANDUM filed by Plaintiff Natera, Inc re 15 MOTION to Seal unredacted versions of the Declaration of Dr. Michael L. Metzker, the Declaration of Solomon Moshkevich, and the entirety of Exhibit F attached to the Moshkevich Declaration [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] filed by Natera, Inc. (VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	17	SEALED UNREDACTED DOCUMENTS filed by Plaintiff Natera, Inc, Defendant Neogenomics Laboratories, Inc. re 13 Declaration, 15 Motion to Seal. (Attachments: # 1 Exhibit 1 - CV, # 2 Exhibit 2 - Kirkizlar et al, # 3 Exhibit 3 - Leon1997Paper, # 4 Exhibit 4 - RaDaR Technology, # 5 Exhibit 5 - Gale, # 6 Exhibit 6 - Analytical-performance, # 7 Exhibit 7 - Forshew, # 8 Exhibit 8 - Heider AACR, # 9 Exhibit 9 - Lipsyc Sharf 2022, # 10 Exhibit 10 - Flach + Supp Methods, # 11 Exhibit 11 - Marsico AACR_2020, # 12 Exhibit 12 - 536-Cutts-AACR-2021, # 13 Exhibit 13 - Plagnol2018Paper, # 14 Exhibit 14 - AACR_2023_Evaluation, # 15 Exhibit 15 - RaDaR_PatientDiscussionGuide, # 16 Exhibit 16 - RaDaR Assay Request Form, # 17 Exhibit 17 - Oncology-Clinical, # 18 Exhibit 18 - jamaoncology, # 19 Exhibit 19 - Coombes2019Paper, # 20 Exhibit 20 - Abbosh2017Paper, # 21 Exhibit 21 - Christensen2019Paper, # 22 Exhibit 22 - Kotani2023Paper) (VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023	18	SEALED UNREDACTED DOCUMENTS filed by Plaintiff Natera, Inc, Defendant Neogenomics Laboratories, Inc. re 15 Motion to Seal, 14 Declaration. (Attachments: # 1 Exhibit A - Natera Receives Final Medicare Coverage, # 2 Exhibit B - Natera Announcement, # 3 Exhibit C - 2022 10-K Natera, # 4 Exhibit D - NeoGenomics Announces Commercial Availability of the RaDaR, # 5 Exhibit E - Neogenomics Investor Day Presentation, # 6 Exhibit F - Email (full document), # 7 Exhibit G - Power Point, # 8 Exhibit H - Refinitiv StreetEvents, # 9 Exhibit I - Refinitiv StreetEvents, # 10 Exhibit J - Russell - AACR 2023, # 11 Exhibit K - 2023_Validation, # 12 Exhibit L - Abbosh 2017, # 13 Exhibit M - Lebow 2022, # 14 Exhibit N - NSCLC clinical data) (VAN ARNAM, ROBERT) (Entered: 07/31/2023)
07/31/2023		Case ASSIGNED to CHIEF JUDGE CATHERINE C. EAGLES and MAG/JUDGE JOE L. WEBSTER. (Bowers, Alexis) (Entered: 08/02/2023)
08/01/2023	<u>19</u>	MOTION for Discovery <i>in advance of Preliminary Injunction Hearing</i> by NATERA, INC Response to Motion due by 8/15/2023 (Attachments: # 1 Text of Proposed Order) (VAN ARNAM, ROBERT) (Entered: 08/01/2023)
08/01/2023	20	MEMORANDUM re 19 MOTION for Discovery in advance of Preliminary Injunction Hearing by Plaintiff NATERA, INC (Attachments: # 1 Exhibit 1. Requests for Production, # 2 Exhibit 2. Interrogatories, # 3 Exhibit 3. 30(b)(6) Depo Notice)(VAN ARNAM, ROBERT) (Entered: 08/01/2023)
08/02/2023		CASE REFERRED to Mediation pursuant to Local Rule 83.9b of the Rules of Practice and Procedure of this Court. Please go to our website under Attorney Information for a list of mediators which must be served on all parties. (Bowers, Alexis) (Entered: 08/02/2023)
08/02/2023	21	Summons Issued as to NEOGENOMICS LABORATORIES, INC (Bowers, Alexis) (Entered: 08/02/2023)

7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Pregre: 264 Filed: 03/18/2024
08/02/2023	22	Notice of Right to Consent. Counsel shall serve the attached form on all parties. (Attachments: # 1 Consent Form) (Bowers, Alexis) (Entered: 08/02/2023)
08/03/2023	24	NOTICE of Attorney Appearance by attorney JOHN F. MORROW, JR on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (MORROW, JOHN) (Entered: 08/03/2023)
08/03/2023	<u>25</u>	Corporate Disclosure Statement identifying Corporate Parent NeoGenomics, Inc. for NEOGENOMICS LABORATORIES, INC (MORROW, JOHN) (Entered: 08/03/2023)
08/03/2023	26	NOTICE by NEOGENOMICS LABORATORIES, INC. re 19 MOTION for Discovery in advance of Preliminary Injunction Hearing [NOTICE OF INTENT TO RESPOND] (MORROW, JOHN) (Entered: 08/03/2023)
08/03/2023	<u>27</u>	STANDARD ORDER for civil cases proceeding before CHIEF JUDGE CATHERINE C. EAGLES.(Winchester, Robin) (Entered: 08/03/2023)
08/04/2023	28	ORDER signed by CHIEF JUDGE CATHERINE C. EAGLES on 08/04/2023. This matter is set for status conference and hearing on the plaintiff's motion for expedited discovery on Wednesday, August 16, 2023, at 11:00 a.m. in the Preyer Courthouse, Courtroom 3, Greensboro, North Carolina. Counsel shall immediately meet and confer about expedited discovery, sealing, scheduling, and any other case management issues that will facilitate prompt, fair, and efficient resolution of pending motions and the case. If the parties reach agreement, they may submit a joint motion and proposed order. (Taylor, Abby) (Entered: 08/04/2023)
08/04/2023		Set Status Conference for 8/16/2023 11:00 AM in Greensboro Courtroom #3 before CHIEF JUDGE CATHERINE C. EAGLES. (Taylor, Abby) (Entered: 08/04/2023)
08/09/2023	29	NOTICE of Special Appearance by attorney EDWARD R. REINES on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3627082.) (REINES, EDWARD) (Entered: 08/09/2023)
08/10/2023	30	NOTICE of Attorney Appearance by attorney CARMELLE F. ALIPIO on behalf of Plaintiff NATERA, INC. (ALIPIO, CARMELLE) (Entered: 08/10/2023)
08/10/2023	31	SUMMONS Returned Executed by NATERA, INC. as to NEOGENOMICS LABORATORIES, INC. served on 8/2/2023, answer due 8/23/2023. (VAN ARNAM, ROBERT) (Entered: 08/10/2023)
08/14/2023	32	RESPONSE re 28 Order Setting Hearing on Motion,, 19 MOTION for Discovery in advance of Preliminary Injunction Hearing [Response to Plaintiff Natera's Motion for Expedited Discovery and The Court's August 4, 2023 Order] filed by NEOGENOMICS LABORATORIES, INC Replies due by 8/28/2023 (Attachments: # 1 Exhibit A E-mail Reines to Haberny 8.11.2023, # 2 Exhibit B - NeoGenomics Discovery Responses, # 3 Exhibit C - Natera Discovery Responses, # 4 Exhibit D - E-mail Reines to Haberny 8.14.2023)(MORROW, JOHN) (Entered: 08/14/2023)
08/15/2023	33	NOTICE of Special Appearance by attorney SANDRA HABERNY on behalf of Plaintiff NATERA, INC. (Filing fee \$ 25 receipt number ANCMDC-3630760.) (HABERNY, SANDRA) (Entered: 08/15/2023)
08/15/2023	34	NOTICE of Special Appearance by attorney VICTORIA MAROULIS on behalf of Plaintiff NATERA, INC. (Filing fee \$ 25 receipt number ANCMDC-3631102.) (MAROULIS, VICTORIA) (Entered: 08/15/2023)
08/15/2023	35	NOTICE of Special Appearance by attorney DEREK WALTER on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-

/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1
08/15/2023	36	REPLY, filed by Plaintiff NATERA, INC., to Response to 19 MOTION for Discovery <i>in advance of Preliminary Injunction Hearing</i> filed by NATERA, INC (VAN ARNAM, ROBERT) (Entered: 08/15/2023)
08/16/2023		Minute Entry for proceedings held before CHIEF JUDGE CATHERINE C. EAGLES: Scheduling Conference held on 8/16/2023. Attorneys Victoria Moroulis, Sandra Haberny, Andrew Shores and Robert Arnman present as counsel for the plaintiff. Attorneys Lisa Ryan, Ed Reines and John Morrow present as counsel for the defendant. Hearing on Motion for Preliminary Injunction set for 11/9/2023 at 9:30 a.m., last brief to be filed on 11/1/2023. The parties are to meet and confer and submit a scheduling order by 8/21/2023. If the parties cannot agree they shall submit a Joint Submission. Any discovery disputes the parties are unable to resolve must be submitted by Joint Submission. (Court Reporter Joseph Armstrong.) (Winchester, Robin) (Entered: 08/16/2023)
08/16/2023		<b>TEXT ORDER</b> issued by CHIEF JUDGE CATHERINE C. EAGLES on 8/16/2023. The parties shall consult regarding Mediator and advise the case manager by 9/25/2023 who they have selected as Mediator.(Winchester, Robin) (Entered: 08/16/2023)
08/16/2023	37	NOTICE of Special Appearance by attorney PATRICK ROBERT LYONS on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3632368.) (LYONS, PATRICK) (Entered: 08/16/2023)
08/17/2023		Set Hearing: Hearing on Motion for Preliminary Injunction set for 11/9/2023 09:30 AM in Greensboro Courtroom #3 before CHIEF JUDGE CATHERINE C. EAGLES. (Winchester, Robin) (Entered: 08/17/2023)
08/17/2023	38	NOTICE of Special Appearance by attorney ELIZABETH RYAN on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3633465.) (RYAN, ELIZABETH) (Entered: 08/17/2023)
08/18/2023	40	L.R. 5.5 Report-Joint filed by all parties by NATERA, INC (VAN ARNAM, ROBERT) (Entered: 08/18/2023)
08/21/2023	41	Joint MOTION for Protective Order by NATERA, INC Response to Motion due by 9/5/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 08/21/2023)
08/21/2023	42	NOTICE of Special Appearance by attorney JOHN J NOLAN on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3635411.) (NOLAN, JOHN) (Entered: 08/21/2023)
08/21/2023	43	JOINT STIPULATION & <i>Proposed Order Regarding Briefing and Discovery Schedule</i> filed by all parties. Est. Trial Days: n/a. (VAN ARNAM, ROBERT) (Entered: 08/21/2023)
08/22/2023		Motions Submitted: <u>41</u> Joint MOTION for Protective Order and case referred re: <u>43</u> Joint Stipulation to CHIEF JUDGE CATHERINE C. EAGLES (Winchester, Robin) (Entered: 08/22/2023)
08/22/2023	44	MOTION for Extension of Time to File Answer re 1 Complaint by NEOGENOMICS LABORATORIES, INC. (Attachments: # 1 Text of Proposed Order) (MORROW, JOHN) (Entered: 08/22/2023)
08/23/2023	45	RESPONSE in Opposition re 44 MOTION for Extension of Time to File Answer re 1 Complaint filed by NEOGENOMICS LABORATORIES, INC. filed by NATERA, INC. (Attachments: # 1 Exhibit 1 - August 21-22, 2023 emails) (VAN ARNAM, ROBERT) (Entered: 08/23/2023)

7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 @@@: 263 Filed: 03/18/2024
08/23/2023		Motions Submitted: 44 MOTION for Extension of Time to File Answer re 1 Complaint to CHIEF JUDGE CATHERINE C. EAGLES- (Winchester, Robin) (Entered: 08/23/2023)
08/23/2023	46	JOINT STIPULATION AND ORDER REGARDING BRIEFING AND DISCOVERY SCHEDULE signed by CHIEF JUDGE CATHERINE C. EAGLES on 08/23/2023, as set out herein.(Taylor, Abby) (Entered: 08/23/2023)
08/23/2023	47	<b>STIPULATED PROTECTIVE ORDER</b> signed by CHIEF JUDGE CATHERINE C. EAGLES on 08/23/2023, as set out herein. (Taylor, Abby) (Entered: 08/23/2023)
08/23/2023	48	ORDER signed by CHIEF JUDGE CATHERINE C. EAGLES on 08/23/2023, that the motion, Doc 44 is GRANTED in part as follows: Any response to the complaint consisting of a motion to dismiss pursuant to Rule 12 of the Rules of Civil Procedure SHALL be filed, along with the required briefing, no later than September 5, 2023. Subject to 1(c), no answer need be filed until the motion to dismiss is resolved, though the defendant may also file answer if it wishes. Any brief in opposition SHALL be filed no later than October 12, 2023, and any reply brief in support SHALL be filed no later than 9 a.m. on October 30, 2023. If the defendant does not file a motion to dismiss, it SHALL file answer no later than September 11, 2023.(Taylor, Abby) (Entered: 08/23/2023)
08/24/2023		<b>TEXT ORDER</b> adopting 46 Local Rule 5.5 Report. Issued by CHIEF JUDGE CATHERINE C. EAGLES on 8/24/23. (Winchester, Robin) (Entered: 08/24/2023)
08/24/2023	49	Transcript of Scheduling Conference held on 08/16/2023 before Judge Catherine C. Eagles.
		Court Reporter J. Armstrong. Phone: 336-332-6034. Email: joseph_armstrong@ncmd.uscourts.gov. Transcript may be viewed at the court public terminal or purchased through the Court Reporter before the deadline for Release of Transcript Restriction. After that date it may be obtained through PACER.
		NOTICE RE: REDACTION OF TRANSCRIPTS: The parties have 5 business days to file a Notice of Intent to Request Redaction and 21 calendar days to file a Redaction Request. If no notice is filed, this transcript will be made electronically available to the public without redaction after 90 calendar days. Transcript may be viewed at the court public terminal or purchased through the court reporter before the 90 day deadline. After that date it may be obtained through PACER.
		Redaction Request due 9/18/2023. Redacted Transcript Deadline set for 9/28/2023. Release of Transcript Restriction set for 11/27/2023. (Armstrong, Joe) (Entered: 08/24/2023)
08/31/2023	50	NOTICE of Special Appearance by attorney KEVIN PB JOHNSON on behalf of Plaintiff NATERA, INC. (Filing fee \$ 25 receipt number ANCMDC-3643405.) (JOHNSON, KEVIN) (Entered: 08/31/2023)
09/05/2023	<u>51</u>	MOTION to Dismiss by NEOGENOMICS LABORATORIES, INC Response to Motion due by 10/12/2023 (MORROW, JOHN) Modified on 9/12/2023 to correct response deadline see Doc. 48 (Winchester, Robin). (Entered: 09/05/2023)
09/05/2023	52	Opening BRIEF re 51 MOTION to Dismiss by Defendant NEOGENOMICS LABORATORIES, INC. filed by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Exhibit 1 - US Patent 11,530,454, # 2 Exhibit 2 -Excerpt from the File History of U.S. Patent No. 11,530,454, # 3 Exhibit 3 - Excerpt from the File History of U.S. Patent No. 11,519,035, # 4 Exhibit 4 - US Patent 11,519,035, # 5 Exhibit 5 - Article By Katherine Elena Varley and Robi David Mitra, # 6 Exhibit 6 - Article By Hui-Yun

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1
		Wang, et al., # 7 Exhibit 7 - Declaration of Professor John Quackenbush, Ph.D., # 8 Exhibit 8 - U.S. Patent Application No. 13 300,235, # 9 Exhibit 9 - Brief of Appellee Natera, Inc. from CareDx, Inc. v. Natera, Inc.)(MORROW, JOHN) (Entered: 09/05/2023)
09/05/2023	53	Additional Attachments to Main Document re 52 Brief,,, . (Attachments: # 1 Exhibit 10 - Excerpt from the File History of U.S. Patent No. 11,519,035, # 2 Exhibit 11 - Excerpt from the File History of U.S Patent No. 11,519,035, # 3 Exhibit 12 - Excerpt from the File History of U.S. Patent Application No. 16 818,842, # 4 Exhibit 13 - Excerpt from the File History of U.S. Patent Application No. 17 164,599, # 5 Exhibit 14 - Declarations Filed During the Prosecution of U.S. Patent Application No. 17164,599, # 6 Exhibit 15 - Article by Honghua Li et al. as published in Comparative Genomics, Vol. 2, edited by Nicholas H. Bergman, in Methods in Molecular Biology (2007 Humana Press Inc.)) (MORROW, JOHN) (Entered: 09/05/2023)
09/05/2023	<u>54</u>	DECLARATION filed by Defendant NEOGENOMICS LABORATORIES, INC. re 52 Brief,,, 53 Additional Attachments to Main Document,, 51 MOTION to Dismiss [DECLARATION OF DEREK C. WALTER] filed by NEOGENOMICS LABORATORIES, INC (MORROW, JOHN) (Entered: 09/05/2023)
09/05/2023	55	Joint MOTION to Compel Discovery by Natera (ROGs 3-6) and by NeoGenomics (ROGs 1 & 5) by NATERA, INC Response to Motion due by 9/19/2023 (Attachments: # 1 Exhibit 1. Aug 28 Emails, # 2 Exhibit 2. NeoGenomics First RFPs, # 3 Exhibit 3. NeoGenomics First ROGs, # 4 Exhibit 4. Natera ROGs 4-7, # 5 Exhibit 5. K. Leicht Aug. 30 Emails, # 6 Exhibit 6. K. Leicht Aug. 31 Emails, # 7 Exhibit 7. Natera ROG Responses, # 8 Exhibit 8. Jamal-Hanjani Thesis, # 9 Exhibit 9. K. Leicht Sept. 5 Emails) (VAN ARNAM, ROBERT) (Entered: 09/05/2023)
09/06/2023	<u>56</u>	Additional Attachments to Main Document re <u>51</u> MOTION to Dismiss . (Attachments: # <u>1</u> Text of Proposed Order)(MORROW, JOHN) (Entered: 09/06/2023)
09/06/2023	<u>57</u>	NOTICE by NEOGENOMICS LABORATORIES, INC. re 55 Joint MOTION to Compel Discovery by Natera (ROGs 3-6) and by NeoGenomics (ROGs 1 & 5) [NOTICE OF ADDITIONAL DISCOVERY SERVED BY PLAINTIFF AFTER PARTIES' JOINT MOTION TO COMPEL] (Attachments: # 1 Exhibit A - Additional Discovery) (MORROW, JOHN) (Entered: 09/06/2023)
09/08/2023		Motions Submitted: <u>55</u> Joint MOTION to Compel Discovery <i>by Natera (ROGs 3-6) and by NeoGenomics (ROGs 1 &amp; 5)</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (Winchester, Robin) (Entered: 09/08/2023)
09/11/2023	<u>58</u>	ANSWER to 1 Complaint, with Jury Demand, Counterclaim to Plaintiff Natera, Inc. against by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Exhibit 1 - US Amendment, # 2 Exhibit 2 - European Patent)(MORROW, JOHN) (Entered: 09/11/2023)
09/14/2023	59	MOTION to Compel Discovery seeking responses to Natera's Second and Third Sets of Interrogatories and Requests for Production by NATERA, INC Response to Motion due by 9/28/2023 (Attachments: # 1 Exhibit 1. NeoGenomics Objections to Natera 1st RFPs, # 2 Exhibit 2. Neogenomics Objections to Natera First ROGs, # 3 Exhibit 3. Natera Second Set of RFPs, # 4 Exhibit 4. Natera Third set of RFPs, # 5 Exhibit 5. Natera Third Set of ROGs, # 6 Exhibit 6. Natera Corrected Second Set of RFPs, # 7 Exhibit 7. Natera Corrected Second Set of ROGs, # 8 Exhibit 8. Natera Corrected Third Set of RFPs, # 9 Exhibit 9. Natera Corrected Third Set of ROGs, # 10 Exhibit 10. Correspondence between counsel proposing Sept. 19 deadline)(VAN ARNAM, ROBERT) Modified on 9/15/2023 to correct response deadline. (Taylor, Abby). (Entered: 09/14/2023)
09/17/2023		<b>TEXT ORDER</b> issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 9/17/23. If the parties do not resolve the motion to compel, Doc. <u>59</u> , Defendant shall

/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1
09/18/2023	60	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 09/18/2023, that Natera's Motion to Compel NeoGenomics to supplement and substantively respond to Natera's Interrogatories No. 3- 6 is GRANTED in part and NeoGenomics shall respond to those Interrogatories to the extent stated here by September 22, 2023. If NeoGenomics intends to raise an issue during preliminary injunction briefing, Natera is entitled to have some inkling of that during the expedited discovery process. NeoGenomics must give Natera some idea of the reasons it already knows it will put forth in opposition to the preliminary injunction motion, based on its own evidence, subject to prompt supplementation as discovery from Natera and additional investigation proceed. The Court does not expect NeoGenomics to provide a comprehensive answer good for all time, but it cannot hide its defenses or arguments against the preliminary injunction based on evidence it already has within its own possession. NeoGenomics's Motion to Compel Natera to supplement its responses to NeoGenomics's Interrogatories No. 1 and No. 5 is GRANTED, and Natera shall respond comprehensively to those Interrogatories by September 25, 2023. Natara has not shown that the discovery sought is not proportional to the needs of the case, and the information sought is likely to be relevant. NeoGenomics does not have to prove relevance beyond a reasonable doubt to obtain discovery. (Taylor, Abby) (Entered: 09/18/2023)
09/18/2023	61	NOTICE of Special Appearance by attorney KEVIN ALEXANDER SMITH on behalf of Plaintiff NATERA, INC., Counter Defendant NATERA, INC. (Filing fee \$ 25 receipt number ANCMDC-3654484.) (SMITH, KEVIN) (Entered: 09/18/2023)
09/19/2023	62	WITHDRAWAL of Motion by Plaintiff NATERA, INC. re <u>59</u> MOTION to Compel Discovery seeking responses to Natera's Second and Third Sets of Interrogatories and Requests for Production filed by NATERA, INC. NeoGenomics to respond to the Interrogatories and produce responsive, non-privileged documents by Friday, Sept. 22, 2023. (SHORES, ANDREW) (Entered: 09/19/2023)
09/22/2023	63	NOTICE of Special Appearance by attorney TARA SRINIVASAN on behalf of Plaintiff NATERA, INC. (Filing fee \$ 25 receipt number ANCMDC-3657713.) (SRINIVASAN, TARA) (Entered: 09/22/2023)
09/25/2023	<u>64</u>	Joint MOTION for Extension of Time to Select Mediator by NATERA, INC (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 09/25/2023)
09/26/2023		Motions Submitted: <u>64</u> Joint MOTION for Extension of Time <i>to Select Mediator</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (Winchester, Robin) (Entered: 09/26/2023)
09/26/2023	<u>65</u>	<b>ORDER</b> granting <u>64</u> Motion for Extension of Time. The Parties shall have until October 2, 2023, to file a Notice of Selection of Mediator. In the absence of an agreement, the Court will appoint a mediator. Signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 09/26/2023. (Taylor, Abby) (Entered: 09/26/2023)
09/28/2023	66	MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery by NEOGENOMICS LABORATORIES, INC (MORROW, JOHN) (Entered: 09/28/2023)
09/28/2023	67	Additional Attachments to Main Document re 66 MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery Exhibit A - 2023-09-25 Natera's 1st Supp Resps & Objs to NeoGenomics Rogs Nos 1 and 5.

/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Page: 266 Filed: 03/18/2024 (Attachments: # 1 Exhibit B - 2023-08-14 Natera's Resps & Objs to NeoGenomics 1st Set of Rogs (Nos. 1-5))(MORROW, JOHN) (Entered: 09/28/2023)
09/28/2023	<u>68</u>	Amended MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery by NEOGENOMICS LABORATORIES, INC (MORROW, JOHN) (Entered: 09/28/2023)
09/29/2023	<u>69</u>	Additional Attachments to Main Document re <u>68</u> Amended MOTION for Extension of Time <i>Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery</i> . (Attachments: # <u>1</u> Text of Proposed Order)(MORROW, JOHN) (Entered: 09/29/2023)
09/29/2023		<b>TEXT ORDER</b> issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 9/29/23 Natera shall file a response to Doc. 68 no later than 9:00 a.m. October 2, 2023. (Winchester, Robin) (Entered: 09/29/2023)
09/29/2023		Set/Response Deadline re <u>68</u> Amended MOTION for Extension of Time <i>Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery</i> : Response to Motion due by 10/2/2023. (Taylor, Abby) (Entered: 09/29/2023)
09/30/2023	70	Joint MOTION for Extension of Time to File Answer re <u>58</u> Answer to Complaint, Counterclaim <i>and Affirmative Defenses</i> by NATERA, INC (Attachments: # <u>1</u> Text of Proposed Order)(SHORES, ANDREW) (Entered: 09/30/2023)
10/01/2023	71	RESPONSE in Opposition re 66 MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery filed by NEOGENOMICS LABORATORIES, INC., 68 Amended MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery filed by NEOGENOMICS LABORATORIES, INC. filed by NATERA, INC Replies due by 10/16/2023 (Attachments: # 1 Exhibit 1. 9.28 Emails, # 2 Exhibit 2. 9.26 Emails, # 3 Exhibit 3. Supp. Response to ROG. 2 (Filed Under Seal), # 4 Exhibit 4. 10.1 Email serving Supp Response to Rog. 2, # 5 Exhibit 5. Supp Response to ROG. 5)(VAN ARNAM, ROBERT) (Entered: 10/01/2023)
10/01/2023	72	MOTION to Seal Ex. 3 - Supp. Response to ROG 2 filed in support of Opp. to Motion to Extend PI Schedule [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NATERA, INC Response to Motion due by 10/16/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 10/01/2023)
10/01/2023	73	MEMORANDUM filed by Plaintiff NATERA, INC. re 72 MOTION to Seal Ex. 3 - Supp. Response to ROG 2 filed in support of Opp. to Motion to Extend PI Schedule [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] filed by NATERA, INC. (VAN ARNAM, ROBERT) (Entered: 10/01/2023)
10/01/2023	74	SEALED UNREDACTED DOCUMENTS Ex. 3 to Natera's Opp to Motion to Extend PI Schedule - Supp Response to ROG 2 filed by Plaintiff NATERA, INC., Defendant NEOGENOMICS LABORATORIES, INC. re 71 Response in Opposition to Motion. (VAN ARNAM, ROBERT) (Entered: 10/01/2023)

2/7/24, 12:10 PM	Case: 24-1324 Document: 42-1 Prage: 267 Filed: 03/18/2024
10/02/2023	Motions Submitted: 70 Joint MOTION for Extension of Time to File Answer re 58 Answer to Complaint, Counterclaim <i>and Affirmative Defenses</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (Winchester, Robin) (Entered: 10/02/2023)
10/02/2023	REPLY, filed by Defendant NEOGENOMICS LABORATORIES, INC., to Response to 66 MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery, 68 Amended MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery filed by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Exhibit 1 - (10.01.23 Email), # 2 Exhibit 2 - (09.26.23 Email), # 3 Exhibit 3 - (10.01.23 - Natera Second Supp Resp and Obj to Rog. 5))(MORROW, JOHN) (Entered: 10/02/2023)
10/02/2023	MOTION for Leave to File Sur-Reply to Defendant's Motion for Extension of Time (Docs 66, 68) by NATERA, INC Response to Motion due by 10/23/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 10/02/2023)
10/02/2023	PLAINTIFF NATERAS SUR-REPLY filed by Plaintiff NATERA, INC. re 69 Amended MOTION for Extension of Time re 76 MOTION for Leave to File Sur-Reply to Defendant's Motion for Extension of Time (Docs. 66, 68) filed by NATERA, INC. (VAN ARNAM, ROBERT) Modified on 10/10/2023 to add text to title of document. (Taylor, Abby). (Entered: 10/02/2023)
10/02/2023	78 ORDER ON JOINT MOTION FOR EXTENSION OF TIME FOR NATERA, TO ANSWER OR RESPOND TO NEOGENOMICS' AFFIRMATIVE DEFENSES COUNTERCLAIMS Doc. 70, signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 10/02/2023, that Natera shall have until October 16, 2023, to file answer or other responsive pleading, including a motion pursuant to Fed. R. Civ. P. 12, to NeoGenomics' Affirmative Defenses and Counterclaims. (Taylor, Abby) (Entered: 10/02/2023)
10/02/2023	79 NOTICE OF SELECTION OF MEDIATOR. (SHORES, ANDREW) (Entered: 10/02/2023)
10/03/2023	Motions Submitted: <u>68</u> Amended MOTION for Extension of Time Neogenomics' Time-Sensitive Motion to Briefly Extend Preliminary Injunction Schedule Due to Plaintiff's Failure to Narrow Asserted Claims and Provide Discovery <u>72</u> MOTION to Seal Ex. 3 - Supp. Response to ROG 2 filed in support of Opp. to Motion to Extend PI Schedule and <u>76</u> MOTION for Leave to File Sur-Reply to Defendant's Motion for Extension of Time (Docs. 66, 68), to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES. (Winchester, Robin) (Entered: 10/03/2023)
10/05/2023	80 ORDER Appointing MARK LEHOCKY, as the Mediator pursuant to LR 83.9d(a). Signed by John S. Brubaker, Clerk of Court. (Bond, Melisa) (Entered: 10/05/2023)
10/05/2023	<b>TEXT ORDER</b> issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 10/5/23. The motion to compel, Doc. 68, remains under advisement. The motion to extend deadlines, Doc. 68, is GRANTED as follows: 1) The hearing on November 9 will begin at 2 pm and the parties shall present a non-argumentative tutorial to assist the Court in understanding background scientific and technical matters in preparation for the preliminary injunction hearing. 2) The preliminary injunction hearing is continued to November 27, 2023, at 9:30 a.m. 3) NeoGenomics deadline to respond to Nateras motion for preliminary injunction is extended to 10 a.m. on October 19, 2023. 4) Nateras deadline to file its reply brief is extended to 9.a.m. on November 6, 2023. Written order to follow as time permits. (Winchester, Robin) Modified on 10/5/2023 to correct the tutorial hearing date, which is November 9 at 2 pm (Winchester, Robin). (Entered: 10/05/2023)

/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Prage: 268 Filed: 03/18/2024
10/06/2023		Set Hearings: Tutorial Hearing set for 11/9/2023 02:00 PM in Greensboro Courtroom #3 before CHIEF DISTRICT JUDGE CATHERINE C. EAGLES. Motion for Preliminary Injunction Hearing set for 11/27/2023 09:30 AM in Greensboro Courtroom #3 before CHIEF DISTRICT JUDGE CATHERINE C. EAGLES. See 10/5/23 Text Order (Winchester, Robin) (Entered: 10/06/2023)
10/06/2023		Set Response and Reply deadlines to Doc. <u>5</u> MOTION for Preliminary Injunction pursuant to 10/6/2023 Text Order. Response to Motion due by 10 a.m. on 10/19/2023; Replies due by 9:00 a.m. on 11/6/2023. (Winchester, Robin) (Entered: 10/06/2023)
10/06/2023	81	Defendant's First Amended ANSWER to 1 Complaint, with Jury Demand, Counterclaim and Affirmative Defenses against by NEOGENOMICS LABORATORIES, INC. (REINES, EDWARD) Modified on 10/10/2023 to modify text. (Taylor, Abby). (Entered: 10/06/2023)
10/10/2023	82	<b>ORDER</b> signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 10/10/2023, that the defendant's motion, Doc. 68, is GRANTED in part and DENIED in part as follows: 1. The hearing on November 9 will begin at 2 p.m., and the parties shall present a tutorial to assist the Court in understanding background and context for scientific and technical issues relevant to the preliminary injunction hearing. 2. The preliminary injunction hearing is continued to November 27, 2023, at 9:30 a.m. 3. NeoGenomics's deadline to respond to Natera's motion for a preliminary injunction is extended to 10 a.m. on October 19, 2023. Natera's deadline to file its reply brief is extended to 9 a.m. on November 6, 2023. 4. Natera's motion for leave to file sur-reply, Doc. 76, is GRANTED, and the Clerk shall correct the docket to reflect that the brief at Doc. 77 is a sur-reply directed to the motion at Doc. 68. (Taylor, Abby) (Entered: 10/10/2023)
10/11/2023	83	Consent MOTION for Extension of Time to File Response/Reply as to <u>51</u> MOTION to Dismiss by NATERA, INC (Attachments: # <u>1</u> Text of Proposed Order)(SHORES, ANDREW) (Entered: 10/11/2023)
10/11/2023		Motions Submitted: <u>83</u> Consent MOTION for Extension of Time to File Response/Reply as to <u>51</u> MOTION to Dismiss to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES-(Winchester, Robin) (Entered: 10/11/2023)
10/12/2023	84	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 10/12/2023, that the Motion, Doc. 83, is GRANTED, and it is ORDERED that the Natera has until October 15, 2023, to file a response to Defendant's Motion to Dismiss, and NeoGenomics has until November 2, 2023 to file a reply in support of its Motion to Dismiss. (Taylor, Abby) (Entered: 10/12/2023)
10/14/2023	85	Consent MOTION for Extension of Time for Extension of Briefing Schedule to Respond to Defendant NeoGenomics Laboratories, Inc.'s Second Amended Answer, Affirmative Defenses and Counterclaims by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Text of Proposed Order)(MORROW, JOHN) (Entered: 10/14/2023)
10/15/2023	86	RESPONSE in Opposition re 51 MOTION to Dismiss filed by NEOGENOMICS LABORATORIES, INC., filed by NATERA, INC. Replies due by 11/2/2023. (Attachments: #1 Exhibit 1 US10538814, #2 Exhibit 2 9580751 FH, #3 Exhibit 3 14692703 FH) (SRINIVASAN, TARA) (Entered: 10/15/2023)
10/16/2023		Motions Submitted: <u>85</u> Consent MOTION for Extension of Time for Extension of Briefing Schedule to Respond to Defendant NeoGenomics Laboratories, Inc.'s Second Amended Answer, Affirmative Defenses and Counterclaims to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (Winchester, Robin) (Entered: 10/16/2023)

7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 @@e: 269 Filed: 03/18/2024
10/16/2023	87	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 10/16/2023, that: 1. The Motion, Doc. 85, is GRANTED, 2. NeoGenomics shall file its Second Amended Answer, Affirmative Defenses, and Counterclaims by October 20, 2023; 3. Natera shall file a responsive pleading, including a motion pursuant to Fed. R. Civ. P. 12, to Defendant's Second Amended Answer, Affirmative Defenses, and Counterclaims by October 27, 2023; 4. NeoGenomics shall file its Opposition to Natera's responsive pleading by November 6, 2023, and 5. Natera shall file its Reply to NeoGenomics' Opposition by November 13, 2023. (Garland, Leah) (Entered: 10/17/2023)
10/18/2023	88	NOTICE of Attorney Appearance by attorney JOHN D. WOOTEN, IV on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (WOOTEN, JOHN) (Entered: 10/18/2023)
10/19/2023	89	RESPONSE in Opposition re <u>5</u> MOTION for Preliminary Injunction filed by NEOGENOMICS LABORATORIES, INC Replies due by 11/2/2023 (MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	90	DECLARATION of Derek Walter re 89 Response in Opposition to Motion by Defendant NEOGENOMICS LABORATORIES, INC. (Attachments: # 1 Exhibit 1 - NEOGEN00058925, # 2 Exhibit 2 - 008 NF Rejection, # 3 Exhibit 3 - 008 Amended Claims, # 4 Exhibit 4 - 008 Applicant Argument, # 5 Exhibit 5 - 5-22-2020 008 Interview Summary, # 6 Exhibit 6 - FILED UNDER SEAL SLIPSHEET, # 7 Exhibit 7 - FILED UNDER SEAL SLIPSHEET, # 8 Exhibit 8 - 454 Applicant Remarks Amendment, # 9 Exhibit 9 - (Metzker Tr.) Excerpted, # 10 Exhibit 10 - FILED UNDER SEAL SLIPSHEET, # 11 Exhibit 11 - FILED UNDER SEAL SLIPSHEET, # 12 Exhibit 12 - FILED UNDER SEAL SLIPSHEET, # 13 Exhibit 13 - Natera 1st Supp to Rogs 1, 5, # 14 Exhibit 14 - NAT-NEO-00002768-2801) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	91	Additional Attachments to Main Document re 90 Declaration (Attachments: # 1 Exhibit 15 - Natera 2nd Supp Rog 5, # 2 Exhibit 16 - Natera 3rd Supp Rog 5), # 3 Exhibit 17 - Inivata Compl., # 4 Exhibit 18 - Natera's R&Os to Rogs 3-4, # 5 Exhibit 19 - FILED UNDER SEAL SLIPSHEET) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	92	Additional Attachments to Main Document re 90 Declaration. (Attachments: # 1 Exhibit 20 - (NEOGEN00004200) - TD Cowen 07.19.23) (MORROW, JOHN) Modified on 10/19/2023 to link to correct pleading. (Sheets, Jamie) (Entered: 10/19/2023)
10/19/2023	93	Additional Attachments to Main Document re 90 Declaration. (Attachments: # 1 Exhibit 21 - NEOGEN00018316, # 2 Exhibit 22 - FILED UNDER SEAL SLIPSHEET, # 3 Exhibit 23 - FILED UNDER SEAL SLIPSHEET, # 4 Exhibit 24 - (NEOGEN00062739) - 09.12.2023 Baird Global Healthcare Conference, # 5 Exhibit 25 - (2023-05-08 Archer Tr.)-Excerpted, # 6 Exhibit 26 - (2023-05-09 Archer Tr.)-Excerpted, # 7 Exhibit 27 - (NEOGEN00059551) - Goldman Sachs 6.13.23 Global HC Conf Trans-Natera, # 8 Exhibit 28 - FILED UNDER SEAL SLIPSHEET) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	94	DECLARATION of Vishal Sikri re 89 Response in Opposition to Motion by Defendant NEOGENOMICS LABORATORIES, INC. (Attachments: # 1 Exhibit 1 - (NEOGEN00014489) - AACR_2020, # 2 Exhibit 2 - (AACR Abstract 3097 re Neo RaDaR Poster) NEOGEN00062633, # 3 Exhibit 3 - 5.5.2021 Neo PR re Inivata Acquisition_NEOGEN00059530, # 4 Exhibit 4 - (NEOGEN00062662) 2021-06-18_NeoGenomics_Completes_Inivata_Acquisition_18, # 5 Exhibit 5 - Inivata 5.25.21 PR_NEOGEN00059564, # 6 Exhibit 6 - NEOGEN00008319, # 7 Exhibit 7 - FILED UNDER SEAL SLIPSHEET, # 8 Exhibit 8 - FDA Breakthrough Devices Program_NEOGEN00059537, # 9 Exhibit 9 - (NeoGenomics 3.16.2023

2/7/24, 12:10 PM	Case	e: 24-1324 Document: 42-1
10/19/2023	95	Additional Attachments to Main Document re 94 Declaration. (Attachments: # 1 Exhibit 11 - Piper Sandler 09.25.23_NEOGEN00059645, # 2 Exhibit 12 - Goldman Sachs 6.13.23 Global HC Conf Trans-Natera_NEOGEN00059551, # 3 Exhibit 13 - (NEOGEN00018316) - MS HC Conf. 09.12.23, # 4 Exhibit 14 - (NEOGEN00062648) 9.12.23 Morgan Stanley Report, # 5 Exhibit 15 - Leelink Partners 7.24.23_NEOGEN00059567) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	96	Additional Attachments to Main Document re 94 Declaration. (Attachments: # 1 Exhibit 16 - Flach, Liquid BIOpsy for MiNimal RESidual DiSease Detection-NEOGEN00001018, # 2 Exhibit 17 - (NEOGEN00004586) - 4.5.23 SVB Securities Report, # 3 Exhibit 18 - FILED UNDER SEAL SLIPSHEET, # 4 Exhibit 19 - NEOGEN00009106, # 5 Exhibit 20 - (NEOGEN00001032) - Gale D, et al. Ann Oncol. 2022 500, # 6 Exhibit 21 - (NEOGEN00000991) - Lipsyc-Sharf-et-al-2022-circulating-tumor, # 7 Exhibit 22 - Elliot_NEOGEN00059699, # 8 Exhibit 23 - Cutts_NEOGEN00059698, # 9 Exhibit 24 - (NEOGEN00011154) - Dasari A, et al. Nat Rev Clin Oncol. 2020 17 757-770, # 10 Exhibit 25 - Rebuzzi_NEOGEN00059710, # 11 Exhibit 26 - Tie_Circulating Tumor_NEOGEN00059740, # 12 Exhibit 27 - Tie 2_NEOGEN00059675, # 13 Exhibit 28 - Taieb_NEOGEN00059722, # 14 Exhibit 29 - (NEOGEN00014540) - Tie J, et al. JAMA Oncol. 2019;5(12)1710-1717, # 15 Exhibit 30 - Chen_NEOGEN00059687, # 16 Exhibit 31 - Natera Oncology_NEOGEN00059701, # 17 Exhibit 32 - FILED UNDER SEAL SLIPSHEET, # 18 Exhibit 33 - NEOGEN0003924) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	97	DECLARATION of Brian Van Ness re 89 Response in Opposition to Motion by Defendant NEOGENOMICS LABORATORIES, INC. (Attachments: # 1 Exhibit 1 - Van Ness CV 1-6-2022, # 2 Exhibit 2 - 5952170 Stroun, # 3 Exhibit 3 - US20100041048A1, # 4 Exhibit 4 - 6156504 Gocke, # 5 Exhibit 5 - 008 NF Rejection, # 6 Exhibit 6 - 2023-09-26 Metzker, Michael_Mini, # 7 Exhibit 7 - Quackenbush Personalis, # 8 Exhibit 8 - 008 Claims, # 9 Exhibit 9 - 008 Applicant Argument, # 10 Exhibit 10 - [DI101] Natera's Opening Brief ISO MSJ, # 11 Exhibit 11 - MSJ Section 101 Hearing - Full, # 12 Exhibit 12 - Metzker Illumina, # 13 Exhibit 13 - US20100120038A1 Fluidigm Mir, # 14 Exhibit 14 - Kaper Abstract, # 15 Exhibit 15 - Kaper Abstract, # 16 Exhibit 16 - Han, # 17 Exhibit 17 - Cantor, # 18 Exhibit 18 - Ehrich, # 19 Exhibit 19 - Lo, # 20 Exhibit 20 - Quackenbush CareDx, # 21 Exhibit 21 - Broude) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	98	Additional Attachments to Main Document re 97 Declaration. (Attachments: # 1 Exhibit 22 - Lyamichev, # 2 Exhibit 23 - Shoemaker, # 3 Exhibit 24 - Wang, # 4 Exhibit 25 - Marsh, # 5 Exhibit 26 - Ahmadian, # 6 Exhibit 27 - Nordstrom(1), # 7 Exhibit 28 - Ronaghi, # 8 Exhibit 29 - 2022-11-09 [DI076] Joint Claim Construction Brief, # 9 Exhibit 30 - Jarvie, # 10 Exhibit 31 - Harismendy and Frazer, # 11 Exhibit 32 - Mardis, # 12 Exhibit 33 - Margulies(1), # 13 Exhibit 34 - Natera MTD Illumina(1), # 14 Exhibit 35 - FILED UNDER SEAL SLIPSHEET, # 15 Exhibit 36 - Metzker Qiagen Trial Testimony, # 16 Exhibit 37 - Metzker Qiagen Trial DemonstrativesDDX-003, # 17 Exhibit 38 - Jamal-Hanjani Thesis, # 18 Exhibit 39 - Shendure, # 19 Exhibit 40 - US4683195, # 20 Exhibit 41 - US20100285478A1) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	99	Additional Attachments to Main Document re 97 Declaration. (Attachments: # 1 Exhibit 42 - US9206475, # 2 Exhibit 43 - Chan, # 3 Exhibit 44 - Leary, # 4 Exhibit 45 - 5-22-2020 008 Interview Summary, # 5 Exhibit 46 - Bentley, # 6 Exhibit 47 - 454 Applicant Remarks Amendment, # 7 Exhibit 48 - 454 NOA, # 8 Exhibit 49 - 454 7-7-2020 Non Final Rejection, # 9 Exhibit 50 - 454 Original Claims, # 10 Exhibit 51 - Bashashati and Supp, # 11 Exhibit 52 - Bashashati S2, # 12 Exhibit 53 - Marzese, # 13 Exhibit 54 -

2/7/24, 12:10 PM	Case: 2	24-1324 Document: 42-1 അതുട് മൂപ് Filed: 03/18/2024
	Ex	ashashati Supplemental Material listing) pathsojournals.onlinelibrary.wiley.com, # 14 khibit 55 - (008 Patent), # 15 Exhibit 59 - coi190019supp1_prod - Reinert) (MORROW, DHN) (Entered: 10/19/2023)
10/19/2023	De Pij Pij FI:	ECLARATION of James E. Malackowski re <u>89</u> Response in Opposition to Motion by efendant NEOGENOMICS LABORATORIES, INC. (Attachments: # 1 Exhibit 1 - per Sandler 2023-09-25 Oncologist Survey (Box-ActiveProjects-2023, # 2 Exhibit 2 - per Sandler 2023-09-25 Oncologist Survey (Box-ActiveProjects-2023, # 3 Exhibit 3 - LED UNDER SEAL SLIPSHEET, # 4 Exhibit 4 - FILED UNDER SEAL SLIPSHEET, 5 Exhibit 5 - Order Denying PI_19-cv-03161-YGR) (MORROW, JOHN) (Entered: \(\frac{1}{19}/2023\))
10/19/2023	6 - Ca Or Ne Ne 14 15 Ne La Ex Ca Ne 19 - N	dditional Attachments to Main Document re 100 Declaration. (Attachments: # 1 Exhibit - S&P Capital IQ - NeoGenomics, Inc. Public Company Profile, # 2 Exhibit 7 - S&P apital IQ - NeoGenomics Laboratories, Inc. Private Company Profile, # 3 Exhibit 8 - ne Lab Vital Answers - NeoGenomics, # 4 Exhibit 9 - Investor Relations - eoGenomics_, # 5 Exhibit 10 - Test Menu - NeoGenomics, # 6 Exhibit 11 - eoGenomics Laboratories LinkedIn, # 7 Exhibit 12 - (NEOGEN00014489) - ACR_2020, # 8 Exhibit 13 - 2021-05-05 NeoGenomics to Acquire Inivata, # 9 Exhibit - 2021-06-18 NeoGenomics Completes Inivata Acquisition, # 10 Exhibit 15 - 2021-05-25 Inivata Announces Collaboration with F-star, # 11 Exhibit 16 - eoGenomics RaDaR Minimal Residual Disease (MRD), # 12 Exhibit 17 - 2023-07-27 eoGenomics Announces Medicare Coverage for RaDaR, # 13 Exhibit 18 - 2023-08-08 abPulse - NeoGenomics secures Medicare coverage for breast cancer blood test, # 14 chibit 19 - MolDX Program (Administered by Palmetto GBA), # 15 Exhibit 20 - S&P apital IQ - Natera, Inc. Public Company Profile, # 16 Exhibit 21 - Corporate Overview - netera, # 17 Exhibit 22 - Natera - Our Technology, # 18 Exhibit 23 - Natera - Oncology, # 18 Exhibit 24 - Natera - Hereditary Genetic Testing for Cancer Empower, # 20 Exhibit 25 Natera - Comprehensive Genomic Profiling (CGP) - Altera) MORROW, JOHN) intered: 10/19/2023)
10/19/2023	26 UN Ex SI 11 8 I - F Di Ce Re Ea 39 gro AS Ca Go Fin	dditional Attachments to Main Document re 100 Declaration. (Attachments: # 1 Exhibit 6 - Natera - Signatera Circulating Tumor DNA Blood Test, # 2 Exhibit 27 - FILED NDER SEAL SLIPSHEET, # 3 Exhibit 28 - FILED UNDER SEAL SLIPSHEET, # 4 thibit 29 - FILED UNDER SEAL SLIPSHEET, # 5 Exhibit 30 - FILED UNDER SEAL LIPSHEET, # 6 Exhibit 31 - FILED UNDER SEAL SLIPSHEET, # 7 Exhibit 32 - 2021-18 - Frontiers in Oncology - Circulating Tumor DNA and Minimal Residual Disease, # Exhibit 33 - Springer Nature - Leading Advances in Precision Oncology, # 9 Exhibit 34 Personalized Cancer Monitoring - Invitae, # 10 Exhibit 35 - Definition of MRD - NCI actionary of Cancer Terms - NCI, # 11 Exhibit 36 - 2020-07-15 - MD Anderson Cancer anter - What is minimal residual disease, # 12 Exhibit 37 - Colorectal Cancer Minimal actional Disease, # 13 Exhibit 38 - 2022-05 - FDA - Use of Circulating Tumor DNA for arlyStage Solid Tumor Drug Development - Draft Guidance for Industry, # 14 Exhibit 10 - DeciBio - Industry Snapshot The nascent ctDNA MRD space continues to see rapid bowth, # 15 Exhibit 40 - DeciBio - ctDNA MRD and Monitoring Front and Center at SCO 2022, # 16 Exhibit 41 - Barrons - How Natera Is Defending Its Lead in a \$15B ancer-Testing Market, # 17 Exhibit 42 - 2023-06-13 - FactSet CallStreet - Natera, Inc Poldman Sachs Global Healthcare Conference, # 18 Exhibit 43 - Guardant Reveal - 10 Exhibit 44 - FILED UNDER SEAL LIPSHEET) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	45 # <u>3</u> FI	dditional Attachments to Main Document re 100 Declaration. (Attachments: # 1 Exhibit 5 - Google Patents - US11530454B2, # 2 Exhibit 46 - Google Patents - US11519035B2, 3 Exhibit 2023-05-08 - Trial Transcript Volume A - Natera v. ArcherD, # 4 Exhibit 48 - LED UNDER SEAL SLIPSHEET, # 5 Exhibit 49 - FILED UNDER SEAL LIPSHEET, # 6 Exhibit 50 - 2023-09-15 NeoGenomics Suppl Resps & Objs to Natera's

2/7/24, 12:10 PM	Case	e: 24-1324 Document: 42-1                 ହେଣୁକ: ଜନଥ         Filed: 03/18/2024
		1st Set of Rogs (Nos. 1-3, # 7 Exhibit 51 - 2022-01-11 NeoGenomics Announces RaDaR Assay Receives CE Mark, # 8 Exhibit 52 - FILED UNDER SEAL SLIPSHEET, # 9 Exhibit 53 - 2015-07-01 Natera Form S-1_Public, # 10 Exhibit 54 - 2015-07-01 Natera Form S-1_Public, # 11 Exhibit 55 - 2020-01-14 - PRNewswire - ArcherDX Personalized Cancer Monitoring) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	104	Additional Attachments to Main Document re 100 Declaration. (Attachments: # 1 Exhibit 56 - Law360 - Natera Nets \$19M Verdict In Cancer Test Patent Case, # 2 Exhibit 57 - Natera, Inc. v. Inivata, Inc - Complaint, # 3 Exhibit 58 - 2022-05-31 - Natera to Present New Signatera Clinical Data Across Multiple Cancer Types - 2022 ASCO Meeting, # 4 Exhibit 59 - Natera Submits First PMA Module to the FDA for Signatera, # 5 Exhibit 60 - GenomeWeb - About Us, # 6 Exhibit 61 - 2023-09-12 - AlphaSense - Baird Global Healthcare Conference, # 7 Exhibit 62 - FILED UNDER SEAL SLIPSHEET, # 8 Exhibit 63 - Natera 2022 Annual Report and Form 10-K) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	105	MOTION to Seal [confidential portions of NeoGenomicss Opposition to Nateras Motion for Preliminary Injunction (Opposition), as well as the accompanying Declarations of Vishal Sikri, with attached Exhibits 7, 18, and 32, and James Malackowski, with attached Exhibits 3 and 62, as well as Exhibits 12 and 28 of the Declaration of Derek Walter] [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NEOGENOMICS LABORATORIES, INC Response to Motion due by 11/2/2023 (Attachments: # 1 Text of Proposed Order)(MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	106	MEMORANDUM filed by Defendant NEOGENOMICS LABORATORIES, INC. re 105 MOTION to Seal [confidential portions of NeoGenomicss Opposition to Nateras Motion for Preliminary Injunction (Opposition), as well as the accompanying Declarations of Vishal Sikri, with attached Exhibits 7, 18, and 32, and James Malackowski, with attached Exhibits 3 and 62, as well as Exhibits 12 and 28 of the Declaration of Derek Walter] filed by NEOGENOMICS LABORATORIES, INC. (Attachments: # 1 Affidavit - Declaration of Alicia Olivo) (MORROW, JOHN) (Entered: 10/19/2023)
10/19/2023	107	SEALED UNREDACTED DOCUMENTS [NeoGenomics' Opp to Natera's Motion for PI] filed by Defendant NEOGENOMICS LABORATORIES, INC (MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	108	SEALED UNREDACTED DOCUMENTS filed by Defendant NEOGENOMICS LABORATORIES, INC. re 107 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 1 - NEOGEN00058925, # 2 Exhibit 2 - 008 NF Rejection, # 3 Exhibit 3 - 008 Amended Claims, # 4 Exhibit 4 - 008 Applicant Argument, # 5 Exhibit 5 - 5-22-2020 008 Interview Summary, # 6 Exhibit 6 - (SEALED) (Moshkevich Tr.)-Excerpted, # 7 Exhibit 7 - (SEALED) Malani Tr Excerpted, # 8 Exhibit 8 - (454 Applicant Remarks Amendment), # 9 Exhibit 9 - (Metzker Tr.) Excerpted, # 10 Exhibit 10 - (SEALED) (Zimmermann Tr.) Excerpted, # 11 Exhibit 11 - (SEALED) (NAT-NEO-00888159, # 12 Exhibit 12 - (SEALED) (9-22-2023 email) HC-AEO, # 13 Exhibit 13 - Natera 1st Supp to Rogs 1, 5), # 14 Exhibit 14 - NAT-NEO-00002768-2801), # 15 Exhibit 15 - Natera 2nd Supp Rog 5)(WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	109	SEALED UNREDACTED DOCUMENTS Declaration of Derek Walter (Additional Exhibits) filed by Defendant NEOGENOMICS LABORATORIES, INC. re 108 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 16 - Natera 3rd Supp Rog, # 2 Exhibit 17 - Inivata Compl, # 3 Exhibit 18 - Natera's R&Os to Rogs 3-4, # 4 Exhibit 19

2/7/24, 12:10 PM	Case	e: 24-1324 Document: 42-1 Prage: 273 Filed: 03/18/2024 - (SEALED) (Natera R&Os to Rog 2)_HC-AEO)(WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	110	SEALED UNREDACTED DOCUMENTS Declaration of Derek Walter (Additional Exhibits) filed by Defendant NEOGENOMICS LABORATORIES, INC. re 108 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 20 - (NEOGEN00004200) - TD Cowen 07.19.23)(WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	111	SEALED UNREDACTED DOCUMENTS Declaration of Derek Walter (Additional Exhibits) filed by Defendant NEOGENOMICS LABORATORIES, INC. re 108 SEALED Unredacted Documents (LR 5.4). (Attachments: #1 Exhibit 21 - NEOGEN00018316, #2 Exhibit 22 - (SEALED) (NAT-NEO-00729487), #3 Exhibit 23 - (SEALED) (NAT-NEO-00864963), #4 Exhibit 24 - (NEOGEN00062739) - 09.12.2023 Baird Global Healthcare Conference, #5 Exhibit 25 - (2023-05-08 Archer Tr.)-Excerpted, #6 Exhibit 26 - (2023-05-09 Archer Tr.)-Excerpted, #7 Exhibit 27 - (NEOGEN00059551) - Goldman Sachs 6.13.23 Global HC Conf Trans-Natera, #8 Exhibit 28 - (SEALED) Beitsch 10.16.2023 Letter)(WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	112	SEALED UNREDACTED DOCUMENTS [DECLARATION OF VISHAL SIKRI] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 107 SEALED Unredacted Documents (LR 5.4). (Attachments: #1 Exhibit (NEOGEN00014489) - AACR_2020, #2 Exhibit 2 - (AACR Abstract 3097 re Neo RaDaR Poster) NEOGEN00062633, #3 Exhibit 3 - 5.5.2021 Neo PR re Inivata Acquisition_NEOGEN00059530, #4 Exhibit 4 - (NEOGEN00062662) 2021-06-18_NeoGenomics_Completes_Inivata_Acquisition_180, #5 Exhibit 5 - Inivata 5.25.21 PR_NEOGEN00059564, #6 Exhibit 6 - (NEOGEN00008319), #7 Exhibit 7 - (SEALED) (NEOGEN00059530) (2.26.2021 FDA Letter Granting Breakthrough Device Designation for RaDaR), #8 Exhibit 8 - FDA Breakthrough Devices Program_NEOGEN00059537, #9 Exhibit 9 - (NeoGenomics 3.16.2023 PR)_NEOGEN00004604, #10 Exhibit 10 (NEOGEN00004200) - TD Cowen 07.19.23)(MORROW, JOHN) Modified on 10/19/2023 to correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	113	SEALED UNREDACTED DOCUMENTS [DECLARATION OF VISHAL SIKRI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 112 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 11 - Piper Sandler 09.25.23 NEOGEN00059645, # 2 Exhibit 12 - Goldman Sachs 6.13.23 Global HC Conf Trans-Natera NEOGEN00059551, # 3 Exhibit 13 - (NEOGEN00018316) - MS HC Conf. 09.12.23, # 4 Exhibit 14 - (NEOGEN00062648) 9.12.23 Morgan Stanley Report, # 5 Exhibit 15 - Leelink Partners 7.24.23 NEOGEN00059567, # 6 Exhibit 16 - Flach, Liquid Biopsy for Minimal Residual Disease Detection-NEOGEN00001018, # 7 Exhibit 17 - (NEOGEN00004586) - 4.5.23 SVB Securities Report, # 8 Exhibit 18 - (SEALED) (NEOGEN00005623) - 5.2023 Neo Peer Group Financial Results, # 9 Exhibit 19 - (NEOGEN00009106), # 10 Exhibit 20 - (NEOGEN00001032) - Gale D, et al. Ann Oncol. 2022)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	114	<b>DUPLICATE OF DOC 113 SEALED UNREDACTED DOCUMENTS</b> [DECLARATION OF VISHAL SIKRI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 112 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 11 - Piper Sandler 09.25.23_NEOGEN00059645, # 2 Exhibit 12 - Goldman Sachs 6.13.23 Global HC Conf Trans-Natera_NEOGEN00059551, # 3 Exhibit 13 - (NEOGEN00018316) - MS HC Conf. 09.12.23, # 4 Exhibit 14 - (NEOGEN00018316) - MS HC Conf. 09.12.23, # 5 Exhibit 15 - Leelink Partners

2/7/24, 12:10 PM	Case	e: 24-1324 Document: 42-1 Prege: 274 Filed: 03/18/2024 7.24.23_NEOGEN00059567, # 6 Exhibit 16 - Leelink Partners 7.24.23_NEOGEN00059567, # 7 Exhibit 17 - (NEOGEN00004586) - 4.5.23 SVB Securities Report, # 8 Exhibit 18 - (SEALED) (NEOGEN00005623) - 5.2023 Neo Peer Group Financial Results, # 9 Exhibit 19 - (NEOGEN00009106), # 10 Exhibit 20 - (NEOGEN00001032) - Gale D, et al. Ann Oncol. 2022)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). Modified on 10/20/2023 (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	115	SEALED UNREDACTED DOCUMENTS [DECLARATION OF VISHAL SIKRI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 112 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 21 - (NEOGEN0000991) - Lipsyc-Sharf-et-al-2022-circulating-tumor, # 2 Exhibit 22 - Elliot_NEOGEN00059699, # 3 Exhibit 23 - Cutts_NEOGEN00059698, # 4 Exhibit 24 - 24 (NEOGEN00011154) - Dasari A, et al. Nat Rev Clin Oncol. 2020 17 757-770, # 5 Exhibit 25 - Rebuzzi_NEOGEN00059710, # 6 Exhibit 26 - Tie_Circulating Tumor_NEOGEN000597406 -, # 7 Exhibit 27 - Tie 2_NEOGEN00059675, # 8 Exhibit 28 - aieb_NEOGEN00059722, # 9 Exhibit 29 - (NEOGEN00014540) - Tie J, et al. JAMA Oncol. 2019;5(12)1710-1717, # 10 Exhibit 30 - Chen_NEOGEN00059687, # 11 Exhibit 31 - Natera Oncology_NEOGEN00059701, # 12 Exhibit 32 - (SEALED) Beitsch 10.16.2023 Letter, # 13 Exhibit 33 - (NEOGEN00003924))(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	116	SEALED UNREDACTED DOCUMENTS Declaration of Dr. Brian Van Ness filed by Defendant NEOGENOMICS LABORATORIES, INC. re 107 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 1 - Van Ness CV 1-6-2022, # 2 Exhibit 2 - 5952170 Stroun., # 3 Exhibit 3 - US20100041048A1, # 4 Exhibit 4 - 6156504 Gocke, # 5 Exhibit 5 - 008 NF Rejection, # 6 Exhibit 6 - 2023-09-26 Metzker, Michael Mini, # 7 Exhibit 7 - Quackenbush Personalis, # 8 Exhibit 8 - 008 Claims, # 9 Exhibit 9 - 008 Applicant Argument, # 10 Exhibit 10 - [DI101] Natera's Opening Brief ISO MSJ, # 11 Exhibit 11 - MSJ Section 101 Hearing, # 12 Exhibit 12 - Metzker Illumina, # 13 Exhibit 13 - US20100120038A1 Fluidigm Mir, # 14 Exhibit 14 - Kaper Abstract, # 15 Exhibit 15 - Kaper, # 16 Exhibit 16 - Han, # 17 Exhibit 17 - Cantor, # 18 Exhibit 18 - Ehrich, # 19 Exhibit 19 - Lo, # 20 Exhibit 20 - Quackenbush CareDx)(WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	117	SEALED UNREDACTED DOCUMENTS Declaration of Dr. Brian Van Ness (Additional Exhibits) filed by Defendant NEOGENOMICS LABORATORIES, INC. re 116 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 21 - Broude, # 2 Exhibit 22 - Lyamichev, # 3 Exhibit 23- Shoemaker, # 4 Exhibit 24 - Wang, # 5 Exhibit 25 - Marsh, # 6 Exhibit 26 - Ahmadian, # 7 Exhibit 27 - Nordstrom, # 8 Exhibit 28 - Ronaghi, # 9 Exhibit 29 - 2022-11-09 [DI076] Joint Claim Construction Brief, # 10 Exhibit 30 - Jarvie, # 11 Exhibit 31 - Harismendy and Frazer, # 12 Exhibit 32 - Mardis, # 13 Exhibit 33 - Margulies, # 14 Exhibit 34 - Natera MTD Illumina, # 15 Exhibit 35 - (SEALED) 2023-10-04 Zimmermann, Bernhard Mini, # 16 Exhibit 36 - Metzker Qiagen Trial Testimony, # 17 Exhibit 37 - Metzker Qiagen Trial DemonstrativesDDX-003) (WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	118	SEALED UNREDACTED DOCUMENTS Declaration of Dr. Brian Van Ness (Additional Exhibits) filed by Defendant NEOGENOMICS LABORATORIES, INC. re 116 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 38 - Jamal-Hanjani Thesis, # 2 Exhibit 39 - Shendure, # 3 Exhibit 40 - US4683195-1, # 4 Exhibit 41 - US20100285478A1, # 5 Exhibit 42 - 42 US9206475., # 6 Exhibit 43 - Chan, # 7 Exhibit 44 - Leary, # 8 Exhibit 45 - 5-22-2020 008 Interview Summary, # 9 Exhibit 46 - Bentley,

2/7/24, 12:10 PM	Case	e: 24-1324 Document: 42-1 Prege: 275 Filed: 03/18/2024  # 10 Exhibit 47 - 454 Applicant Remarks Amendment, # 11 Exhibit 48 - 454 NOA, # 12 Exhibit 49 - 454 7-7-2020 Non Final Rejection, # 13 Exhibit 50 - 454 Original Claims, # 14 Exhibit 51 - Bashashati and Supp, # 15 Exhibit 52 - Bashashati S2, # 16 Exhibit 53 - Marzese, # 17 Exhibit 54 - (Bashashati Supplemental Material listing) pathsojournals.onlinelibrary.wiley.com, # 18 Exhibit 55 - 008 Patent, # 19 Exhibit 59 - coi190019supp1_prod - Reinert)(WOOTEN, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	119	SEALED UNREDACTED DOCUMENTS [DECLARATION OF JAMES E. MALACKOWSKI] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 107 SEALED Unredacted Documents (LR 5.4). (Attachments: #1 Exhibit 1 - Piper Sandler 2023-09-25 Oncologist Survey (Box-ActiveProjects-2023, #2 Exhibit 2 - (NEOGEN00018316) - MS HC Conf. 09.12.23, #3 Exhibit 3 - (SEALED) AEO_NeoGenomics - Supp. Resp to Rogs 3 & 6, #4 Exhibit 4 - (SEALED) AEO_NAT-NEO-00584060, #5 Exhibit 5 - Order Denying PI_19-cv-03161-YGR, #6 Exhibit 6 - S&P Capital IQ - NeoGenomics, Inc. Public Company Profile_, #8 Exhibit 8 - One Lab Vital Answers - NeoGenomics, Inc. Private Company Profile_, #8 Exhibit 8 - One Lab Vital Answers - NeoGenomics, #9 Exhibit 9 - Investor Relations - NeoGenomics_, #10 Exhibit 10 - Test Menu - NeoGenomics, #11 Exhibit 11 - NeoGenomics Laboratories LinkedIn, #12 Exhibit 12 - (NEOGEN00014489) - AACR_2020, #13 Exhibit 13 - 2021-05-05 NeoGenomics to Acquire Inivata, #14 Exhibit 14 - 2021-06-18 NeoGenomics Completes Inivata Acquisition, #15 Exhibit 15 - 2021-05-25 Inivata Announces Collaboration with F-star, #16 Exhibit 16 - NeoGenomics RaDaR Minimal Residual Disease (MRD), #17 Exhibit 17 - 2023-07-27 NeoGenomics Announces Medicare Coverage for RaDaR, #18 Exhibit 18 - 2023-08-08 LabPulse - NeoGenomics secures Medicare coverage for breast cancer blood test, #19 Exhibit 19 - MoIDX Program (Administered by Palmetto GBA), #20 Exhibit 20 - S&P Capital IQ - Natera, Inc. Public Company Profile_)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	120	SEALED UNREDACTED DOCUMENTS [DECLARATION OF JAMES E. MALACKOWSKI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 119 SEALED Unredacted Documents (LR 5.4). (Attachments: #1 Exhibit 21 - Corporate Overview - Natera, #2 Exhibit 22 - Natera - Our Technology, #3 Exhibit 23 - Natera - Oncology, #4 Exhibit 24 - Hereditary Genetic Testing for Cancer Empower, #5 Exhibit 25 - Natera - Comprehensive Genomic Profiling, #6 Exhibit 26 - Natera - Signatera Circulating Tumor DNA Blood Test, #7 Exhibit 27 - (SEALED) AEO_NAT-NEO-00612643, #8 Exhibit 28 - (SEALED) AEO_NAT-NEO-00660081, #9 Exhibit 29 - (SEALED) AEO_NAT-NEO-00587463, #10 Exhibit 30 - (SEALED) AEO_NAT-NEO-00590295, #11 Exhibit 31 - (SEALED) AEO_Moshkevich 2023-09-22, #12 Exhibit 32 - 2021-11-18 - Frontiers in Oncology - Circulating Tumor DNA and Minimal Residual Disease, #13 Exhibit 33 - Springer Nature - Leading Advances in Precision Oncology, #14 Exhibit 34 - Personalized Cancer Monitoring - Invitae, #15 Exhibit 35 - Definition of MRD - NCI Dictionary of Cancer Terms - NCI, #16 Exhibit 36 - 2020-07-15 - MD Anderson Cancer Center - What is minimal residual disease, #17 Exhibit 37 - Colorectal Cancer Minimal Residual Disease_, #18 Exhibit 38 - 2022-05 - FDA - Use of Circulating Tumor DNA for Early Stage Solid Tumor Drug Development - Draft Guidance for Industry, #19 Exhibit 39 - DeciBio - Industry Snapshot The nascent ctDNA MRD space continues to see rapid growth)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	121	SEALED UNREDACTED DOCUMENTS [DECLARATION OF JAMES E. MALACKOWSKI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 119 SEALED Unredacted Documents (LR 5.4).

2/7/24, 12:10 PM	122	e: 24-1324 Document: 42-1 Prege: 276 Filed: 03/18/2024  (Attachments: # 1 Exhibit 40 - DeciBio - ctDNA MRD and Monitoring Front and Center at ASCO 2022, # 2 Exhibit 41 - Barrons - How Natera Is Defending Its Lead in a \$15B Cancer-Testing Market, # 3 Exhibit 42 - 2023-06-13 - FactSet CallStreet - Natera, Inc Goldman Sachs Global Healthcare Conference Public, # 4 Exhibit 43 - DeciBio - ctDNA MRD and Monitoring Front and Center at ASCO 2022)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)  SEALED UNREDACTED DOCUMENTS [DECLARATION OF JAMES E. MALACKOWSKI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 119 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 44 - (SEALED) AEO Malani 2023-10-02 Mini Depo, # 2 Exhibit 45 - Google Patents - US11530454B2)(MORROW, JOHN) Modified on 10/10/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/10/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/10/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/10/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/10/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/10/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/10/2023 to reflect correct filer of document (Winchester, Robin).
10/19/2023	123	10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)  SEALED UNREDACTED DOCUMENTS [DECLARATION OF JAMES E. MALACKOWSKI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 119 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 46 - Google Patents - US11519035B2, # 2 Exhibit 47 - 023-05-08 - Trial Transcript Volume A - Natera v. ArcherDX_, # 3 Exhibit 48 - (SEALED) AEO_Moshkevich 2023-09-22 Ex. 4-, # 4 Exhibit 49 - (SEALED) AEO_Moshkevich 2023-09-22 Ex. 3, # 5 Exhibit 50 - 2023-09-15 NeoGenomics Suppl Resps & Objs to Natera's 1st Set of Rogs (Nos. 1-3), # 6 Exhibit 51 - 2022-01-11 NeoGenomics Announces RaDaR Assay Receives CE Mark, # 7 Exhibit 52 - (SEALED) AEO_NAT-NEO-00729334, # 8 Exhibit 53 - 2015-07-01 Natera Form S-1, # 9 Exhibit 54 - Natera, Inc. v. ArcherDX, Inc Verdict Form, # 10 Exhibit 55 - 020-01-14 - PRNewswire - ArcherDX Personalized Cancer Monitoring)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	124	SEALED UNREDACTED DOCUMENTS [DECLARATION OF JAMES E. MALACKOWSKI (additional documents)] filed by Defendant NEOGENOMICS LABORATORIES, INC. re 119 SEALED Unredacted Documents (LR 5.4). (Attachments: # 1 Exhibit 56 - Law360 - Natera Nets \$19M Verdict In Cancer Test Patent Case_, # 2 Exhibit 57 - Natera, Inc. v. Inivata, Inc - Complaint, # 3 Exhibit 58 - 2022-05-31 - Natera to Present New Signatera Clinical Data Across Multiple Cancer Types - 2022 ASCO Meeting, # 4 Exhibit 59 - 2023-10-02 - Natera Submits First PMA Module to the FDA for Signatera, # 5 Exhibit 60 - GenomeWeb - About Us, # 6 Exhibit 61 - 2023-09-12 - AlphaSense - Baird Global Healthcare Conference, # 7 Exhibit 62 - (SEALED) AEO_2023-10-16 Beitsch Letter, # 8 Exhibit 63 - Natera 2022 Annual Report and Form 10-K)(MORROW, JOHN) Modified on 10/19/2023 to reflect correct filer of document (Winchester, Robin). (Entered: 10/19/2023)
10/19/2023	125	NOTICE of Special Appearance by attorney JOSHUA HIRAM HARRIS on behalf of Plaintiff NATERA, INC. (Filing fee \$ 25 receipt number BNCMDC-3677497.) (HARRIS, JOSHUA) (Entered: 10/19/2023)
10/20/2023	126	REDACTED SECOND AMENDED ANSWER to 1 Complaint with Jury Demand, Counterclaim against by NEOGENOMICS LABORATORIES, INC. (REINES, EDWARD) Modified on 10/23/2023 to clean up text. (Taylor, Abby). (Entered: 10/20/2023)
10/20/2023	127	MOTION to Seal Second Amended Answer, Affirmative Defenses & Counterclaims [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NEOGENOMICS LABORATORIES, INC Response to Motion due

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Proposed Order)(REINES, EDWARD) (Entered: 10/20/2023)
10/20/2023	128	Second Amended Answer, Affirmative Defenses & Counterclaims filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC., Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. re 126 Answer to Complaint, Counterclaim. (REINES, EDWARD) Modified on 1/16/2024 to unseal document, see 198 Order. (at). (Entered: 10/20/2023)
10/24/2023	129	NOTICE of Special Appearance by attorney AUGUST MELCHER on behalf of Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3681508.) (MELCHER, AUGUST) (Entered: 10/24/2023)
10/27/2023	130	MOTION TO DISMISS FOR FAILURE TO STATE A CLAIM and Motion to Strike Affirmative Defenses by NATERA, INC Response to Motion due by 11/6/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 10/27/2023)
10/27/2023	131	MEMORANDUM filed by Counter Defendant NATERA, INC. re 130 MOTION TO DISMISS FOR FAILURE TO STATE A CLAIM and Motion to Strike Affirmative Defenses REDACTED filed by NATERA, INC (VAN ARNAM, ROBERT) (Entered: 10/27/2023)
10/27/2023	132	MOTION to Seal <i>Unredacted Memorandum in Support of Motion to Dismiss Fifth Counterclaim and Strike Affirmative Defenses 10 &amp; 11</i> [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NATERA, INC (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 10/27/2023)
10/27/2023	133	MEMORANDUM filed by Counter Defendant NATERA, INC. re 127 MOTION to Seal Second Amended Answer, Affirmative Defenses & Counterclaims [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).], 132 MOTION to Seal Unredacted Memorandum in Support of Motion to Dismiss Fifth Counterclaim and Strike Affirmative Defenses 10 & 11 [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).], filed by NATERA, INC. (VAN ARNAM, ROBERT) (Entered: 10/27/2023)
10/27/2023	134	Unredacted Memo in Support of Motion to Dismiss Fifth Counterclaim and Motion to Strike Tenth and Eleventh Affirmative Defenses filed by Counter Defendant NATERA, INC., Plaintiff NATERA, INC., Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimant NEOGENOMICS LABORATORIES, INC. re 131 Memorandum. (VAN ARNAM, ROBERT) Modified on 1/16/2024 to unseal document, see 198 Order. (at). (Entered: 10/27/2023)
11/02/2023	135	REPLY in Support re <u>51</u> MOTION to Dismiss filed by NEOGENOMICS LABORATORIES, INC Replies due by 11/16/2023 (MORROW, JOHN) Modified on 11/3/2023 to correct title of document (Winchester, Robin). (Entered: 11/02/2023)
11/02/2023	<u>136</u>	Unopposed Motion for Lead Counsel to Remotely Attend Tutorial Hearing by NEOGENOMICS LABORATORIES, INC. (MORROW, JOHN) (Entered: 11/02/2023)

/24, 12:10 PM	Case: 24-1324 Document: 42-1 পেল্ডেছ: প্রনষ্ট Filed: 03/18/2024
11/03/2023	Motions Submitted: <u>136</u> <i>Unopposed Motion for Lead Counsel to Remotely Attend Tutorial Hearing</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES (Winchester Robin) (Entered: 11/03/2023)
11/03/2023	<b>TEXT ORDER</b> issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 11/3/23. The motion, Doc. 136, is GRANTED to this extent: The Clerk will make arrangements for Mr. Reines to attend the hearing remotely, with such audio and video access as the Clerk can easily arrange. But the Court will only hear from counsel who are present in person, which works better logistically and substantively than mixed inperson/remote hearings. There are plenty of capable attorneys for the defendant and this hearing date has been set aside for weeks. (Winchester, Robin) (Entered: 11/03/2023)
11/03/2023	Motions Submitted: 5 MOTION for Preliminary Injunction; 15 MOTION to Seal unredacted versions of the Declaration of Dr. Michael L. Metzker, the Declaration of Solomon Moshkevich, and the entirety of Exhibit F attached to the Moshkevich Declaration; 51 MOTION to Dismiss; 72 MOTION to Seal Ex. 3 - Supp. Response to ROG 2 filed in support of Opp. to Motion to Extend PI Schedule; 105 Motion to Seal [confidential portions of NeoGenomicss Opposition to Nateras Motion for Preliminary Injunction (Opposition), as well as the accompanying Declarations of Vishal Sikri, with attached Exhibits 7, 18, and 32, and James Malackowski to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (Winchester, Robin) (Entered: 11/03/2023)
11/03/2023	MOTION to Bind Natera to its Asserted Conception Date in its Court-Ordered Discover Response by NEOGENOMICS LABORATORIES, INC. Response to Motion due by 11/24/2023 (Attachments: # 1 Text of Proposed Order Proposed Order)(MORROW, JOHN) (Entered: 11/03/2023)
11/03/2023	MEMORANDUM filed by Defendant NEOGENOMICS LABORATORIES, INC. re 13 MOTION to Bind Natera to its Asserted Conception Date in its Court-Ordered Discover Response filed by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Exhibit 1- (2023-09-25) Natera's 1st Supp Resps & Objs to NeoGenomics Rogs Nos. 1 and 5, # Exhibit 2 - (2023-10-26) Natera's 2nd Supp Resp & Objs to Rog No. 1, # 3 Exhibit 3 - (2023-09-15) NAT-NEO-00534995-NAT-NEO-00883917, # 4 Exhibit 4 - (2023-09-21) NAT-NEO-00884055-NAT-NEO-00887706)(MORROW, JOHN) (Entered: 11/03/2023)
11/06/2023	REPLY, filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC., to Response to 5 MOTION for Preliminary Injunction Redact filed by NATERA, INC (Attachments: # 1 Exhibit 1 - FILED UNDER SEAL, # 2 Exhibit 2 - FILED UNDER SEAL, # 3 Exhibit 3 - FILED UNDER SEAL, # 4 Exhibit 4 FILED UNDER SEAL - Zimmermann Depo Tr, # 5 Exhibit 5 - FILED UNDER SEAL Malackowski Depo Tr, # 6 Exhibit 6 - FILED UNDER SEAL - Discovery Responses, # Exhibit 7 - FILED UNDER SEAL - Sikri Depo Tr, # 8 Exhibit 8 - FILED UNDER SEAL - Email)(VAN ARNAM, ROBERT) (Entered: 11/06/2023)
11/06/2023	DECLARATION filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC. re 139 Reply to Response to Motion, of Rob L. Stoll filed by NATERA, INC (Attachments: # 1 Exhibit A - CV, # 2 Exhibit B - Supplemental Application Data Sheet)(VAN ARNAM, ROBERT) (Entered: 11/06/2023)
11/06/2023	DECLARATION filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC. re 139 Reply to Response to Motion, OF DE MICHAEL METZKER filed by NATERA, INC. (Attachments: # 1 Exhibit 1 - FILED UNDER SEAL - Transcript of Brian VanNess, # 2 Exhibit 2 - Cell-free DNA challenges 3 Exhibit 3 - Field guide to next-generation DNA sequencers, # 4 Exhibit 4 - Detection ultra-rare mutations by next-generation sequencing, # 5 Exhibit 5 - Kinde2011Paper, # 6 Exhibit 6 - Metzker2006PCRChapter, # 7 Exhibit 7 - Gilles2011Paper, # 8 Exhibit 8 -

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Page: 279 Filed: 03/18/2024  Metzker2010Review, # 9 Exhibit 9 - Kuleshov2014Paper, # 10 Exhibit 10 - GenomeWeb, # 11 Exhibit 11 - Protocol, # 12 Exhibit 12 - NeoGenomics_Announces_Medicare_Coverage)(VAN ARNAM, ROBERT) (Entered: 11/06/2023)
11/06/2023	142	MOTION to Seal Reply in Support of its Motion for Preliminary Injunction Exhibits 1-8 attached to Nateras Reply, portions of the Reply Declaration of Dr. Michael L. Metzker, Ph.D, and Exhibit 1 attached to the Metzker Declaration [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NATERA, INC Response to Motion due by 11/20/2023 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 11/06/2023)
11/06/2023	143	MEMORANDUM filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC. re 142 MOTION to Seal Reply in Support of its Motion for Preliminary Injunction Exhibits 1-8 attached to Nateras Reply, portions of the Reply Declaration of Dr. Michael L. Metzker, Ph.D, and Exhibit 1 attached to the Metzker Declaration filed by NATERA, INC (VAN ARNAM, ROBERT) (Entered: 11/06/2023)
11/06/2023	144	SEALED UNREDACTED DOCUMENTS filed by Plaintiff NATERA, INC., Defendant NEOGENOMICS LABORATORIES, INC. re 139 Reply to Response to Motion. (Attachments: #1 Exhibit 1 - NAT-NEO-00884440, #2 Exhibit 2 - NAT-NEO-00574046, #3 Exhibit 3 - NAT-NEO-00774804, #4 Exhibit 4 - Zimmermann Depo Tr, #5 Exhibit 5 - Malackowski Depo Tr, #6 Exhibit 6 - Discovery Responses, #7 Exhibit 7 - Sikri Depo Tr, #8 Exhibit 8 - Email) (VAN ARNAM, ROBERT) Modified on 1/16/2024 to useal Exhibit 8, see 198 Order. (at). (Entered: 11/06/2023)
11/06/2023	145	SEALED UNREDACTED DOCUMENTS filed by Plaintiff NATERA, INC., Defendant NEOGENOMICS LABORATORIES, INC. re 141 Declaration. (Attachments: # 1 Exhibit 1 - Transcript of BrianVanNess) (VAN ARNAM, ROBERT) (Entered: 11/06/2023)
11/06/2023	146	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 11/06/2023. To assist the Court in evaluating the pending motions to seal and in locating evidence easily, the parties SHALL work together cooperatively to prepare and the plaintiff SHALL file on the public docket, no later than November 30, 2023, a Joint Submission in aid of the pending motions to seal. The Joint Submission SHALL contain the following information in chart form, organized by the substantive motion at issue and listing each document subject to a pending motion to seal: The name of the document or summary identification of document (i.e., Brief in opposition to motion for preliminary injunction, or 10/16/23 Beitsch Letter); the location of the public redacted version or slip sheet (i.e., Doc. 103-5), the location of the temporarily-sealed unredacted version (i.e., Doc. 111-3); and the docket number of the motion to seal applicable to the document. For a somewhat similar example, see Seaman v. Duke, 15cv462, at Doc. 169. If the parties prefer, they can organize these charts by motion to seal, so long as the chart makes it clear what substantive motion(s) the documents and briefs relate to. If the parties have a better idea on how to present this information, the Court is open to suggestion. (Taylor, Abby) (Entered: 11/06/2023)
11/06/2023	147	RESPONSE in Opposition re 130 MOTION TO DISMISS FOR FAILURE TO STATE A CLAIM and Motion to Strike Affirmative Defenses filed by NATERA, INC. filed by NEOGENOMICS LABORATORIES, INC Replies due by 11/20/2023 (REINES, EDWARD) (Entered: 11/06/2023)
11/09/2023		Minute Entry for proceedings held before CHIEF DISTRICT JUDGE CATHERINE C. EAGLES: Tutorial Hearing held on 11/9/2023. Attorneys Sandara Haberny, Tara Srinivasan, Robert Van Arnman and In House Counsel Sean Boyle present as counsel for plaintiff. Attorneys Elizabeth Ryan, John Morrow appeared in person, Attorney Edward

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Page: 280 Filed: 03/18/2024  Reienes appeaed via VTC as counsel for defendant. Plaintiff shall file response to Doc. 137 on 11/17/2023, no reply allowed. (Court Reporter Briana Chesnut.) (Winchester, Robin) (Entered: 11/09/2023)					
11/13/2023	148	REPLY, filed by Counter Defendants NATERA, INC., NATERA, INC., Plaintiff NATERA, INC., to Response to 130 MOTION TO DISMISS FOR FAILURE TO STATE A CLAIM and Motion to Strike Affirmative Defenses filed by NATERA, INC (VAN ARNAM, ROBERT) (Entered: 11/13/2023)					
11/13/2023	149	Reply Brief in Support of Motion to Dismiss and Strike Affirmative Defenses filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC., Counter Claimant NEOGENOMICS LABORATORIES, INC. re 132 Motion to Seal, 130 Motion to Dismiss for Failure to State a Claim. (VAN ARNAM, ROBERT) Modified on 1/16/2024 to unseal document, see 198 Order. (at). (Entered: 11/13/2023)					
11/14/2023		Motions Submitted: 130 MOTION TO DISMISS FOR FAILURE TO STATE A CLAIM and Motion to Strike Affirmative Defenses, 132 MOTION to Seal Unredacted Memorandum in Support of Motion to Dismiss Fifth Counterclaim and Strike Affirmative Defenses 10 & 11 [If the party filing this motion is not the party claiming confidentiality, the party claiming conf to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (rw) (Entered: 11/14/2023)					
11/14/2023	<u>150</u>	NOTICE by NATERA, INC. regarding witnesses it may call during PI Hearing (SHORES, ANDREW) (Entered: 11/14/2023)					
11/16/2023		<b>TEXT ORDER</b> issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 11/16/23. To assist the Court, each party SHALL file proposed findings of fact and conclusions of law by 8:00 a.m. on December 1, 2023, directed to the likelihood of success on the merits element of a preliminary injunction. They may address the other factors, but that is not required. Each statement of fact should be followed by citation(s) to the docket supporting the finding, using the CM/ECF docket numbering and pagination in the following format: Doc. [ENTRY #] at [PAGE #] or Doc. [Entry #] at paragraph [#], i.e., Doc. 4 at 6 or Doc. 7 at paragraph 8. To the extent a party relies on testimony at the hearing, no cite to a transcript is required but the party SHALL identify the witness and SHALL also cite the place in the record where other factual support is found. Each party SHALL provide a copy in Word format to the case manager. (rw) Modified on 11/16/2023 to reflect the correct date, which is December 1, 2023. (rw). (Entered: 11/16/2023)					
11/17/2023	<u>151</u>	NOTICE by NEOGENOMICS LABORATORIES, INC. regarding witnesses it may call during PI Hearing (REINES, EDWARD) (Entered: 11/17/2023)					
11/17/2023	152	See 155 for corrected Brief RESPONSE in Opposition re 137 MOTION NeoGenomics' Motion to Bind Natera to its Asserted Conception Date in its Court-Ordered Discovery Response filed by NEOGENOMICS LABORATORIES, INC. filed by NATERA, INC Replies due by 11/17/2023 (Attachments: # 1 Exhibit 1. FILED UNDER SEAL, # 2 Exhibit 2. NeoGenomics' 30(b)(6) Depo Notice, # 3 Exhibit 3. 30(b)(6) Witness Designations, # 4 Exhibit 4. Natera Doc Prod Emails, # 5 Exhibit 5. 9.28 Reines Email, # 6 Exhibit 6. 10.16 Harris Email, # 7 Exhibit 7. Zimmermann Decl. (11.17.2023), # 8 Exhibit 8. Srinivasan Email, # 9 Exhibit 9. FILED UNDER SEAL, # 10 Exhibit 10. Srinivasan Decl.)(VAN ARNAM, ROBERT) Modified on 11/20/2023 to reflect corrected brief filed at 155 Response. (km) (Entered: 11/17/2023)					
11/17/2023	153	SEALED UNREDACTED DOCUMENTS Unredacted Opp. to Neo FRCP 37 Motion to Bind Natera to An Incorrect Conception Date, and Unredacted Exhibits 1 and 9 thereto filed by Plaintiff NATERA, INC. re 132 Motion to Seal, 152 Response in Opposition to Motion,, 142 Motion to Seal, (Attachments: # 1 Exhibit 1. FILED UNDER SEAL, # 2					

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Page: 284 Filed: 03/18/2024  Exhibit 2. NeoGenomics' 30(b)(6) Depo Notice, # 3 Exhibit 3. 30(b)(6) Witness Designations, # 4 Exhibit 4. Natera Doc Prod Emails, # 5 Exhibit 5. 9.28 Reines Email, # 6 Exhibit 6. 10.16 Harris Email, # 7 Exhibit 7. Zimmermann Decl. (11.17.2023), # 8 Exhibit 8. Srinivasan Email, # 9 Exhibit 9. FILED UNDER SEAL, # 10 Exhibit 10. Srinivasan Decl.)(VAN ARNAM, ROBERT) Modified on 1/16/2024 to unseal exhibit 1 and exhibit 9, see 198 Order (at). (Entered: 11/17/2023)	
		Transcript of Proceedings held on 11/9/2023, before Judge Catherine C. Eagles. Court Reporter Briana L. Chesnut, Telephone number 336-734-2514. Email: brinesbit@gmail.com. Transcript may be viewed at the court public terminal or purchased through the Court Reporter before the deadline for Release of Transcript Restriction. After that date it may be obtained through PACER.	
		NOTICE RE: REDACTION OF TRANSCRIPTS: The parties have 5 business days to file a Notice of Intent to Request Redaction and 21 calendar days to file a Redaction Request. If no notice is filed, this transcript will be made electronically available to the public without redaction after 90 calendar days. Transcript may be viewed at the court public terminal or purchased through the court reporter before the 90 day deadline. After that date it may be obtained through PACER.	
		Redaction Request due 12/14/2023. Redacted Transcript Deadline set for 12/22/2023. Release of Transcript Restriction set for 2/20/2024. (bc) (Entered: 11/18/2023)	
11/20/2023		Motions Submitted: <u>137</u> MOTION <i>NeoGenomics' Motion to Bind Natera to its Asserted Conception Date in its Court-Ordered Discovery Response</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (rw) (Entered: 11/20/2023)	
11/20/2023	<u>155</u>	Corrected document re <u>152</u> Response in Opposition to Motion. (VAN ARNAM, ROBERT) (Entered: 11/20/2023)	
11/21/2023	156	NOTICE of Special Appearance by attorney ANDREW J BRAMHALL on behalf of Plaintiff NATERA, INC., Counter Defendants NATERA, INC., NATERA, INC. (Filing fee \$ 25 receipt number ANCMDC-3704976.) (BRAMHALL, ANDREW) (Entered: 11/21/2023)	
11/22/2023	157	Suggestion of Subsequently Decided Authority re <u>5</u> MOTION for Preliminary Injun by Plaintiff NATERA, INC (SHORES, ANDREW) (Entered: 11/22/2023)	
11/22/2023	158	NOTICE by NATERA, INC. JOINT Notice regarding anticipated witnesses at PI Hearing (SHORES, ANDREW) (Entered: 11/22/2023)	
11/27/2023	159	Minute Entry for proceedings held before CHIEF DISTRICT JUDGE CATHERINE C. EAGLES. Attorneys Kevin Johnson, Victoria Maroulis, Andrew Bramhall, Sandra Haberny, Tara Srinivasan and Robert Van Arnam present as counsel for plaintiff. Attorneys John Morrow, Jr., Derek Walter, and Edward Reines present as counsel for the defendant. Motion Hearing held on 11/27/2023 re 5 MOTION for Preliminary Injunction filed by NATERA, INC. Evidence presented see exhibit list. Matter taken under advisement. (Court Reporter Joseph Armstrong.) (rw) (Entered: 11/27/2023)	
11/29/2023	160	RESPONSE in Support re 105 MOTION to Seal [confidential portions of NeoGenomicss Opposition to Nateras Motion for Preliminary Injunction (Opposition), as well as the accompanying Declarations of Vishal Sikri, with attached Exhibits 7, 18, and 32, and James Malackowski, with filed by NATERA, INC Replies due by 12/13/2023 (VAN ARNAM, ROBERT) (Entered: 11/29/2023)	
11/30/2023	161	JOINT STATUS REPORT regarding Locations of Sealed and Public Documents (pursuant to Order at Doc. 146) filed by all parties. Est. Trial Days: TBD. (VAN	

/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Proge: 282 Filed: 03/18/2024 ARNAM, ROBERT) (Entered: 11/30/2023)					
12/01/2023	162	Statement of Material Facts re <u>5</u> MOTION for Preliminary Injunction [NEOGENOMICS' PROPOSED FINDINGS OF FACT AND CONCLUSIONS OF LAW IN SUPPORT OF ITS OPPOSITION TO NATERA'S MOTION FOR A PRELIMINARY INJUNCTION] filed by NEOGENOMICS LABORATORIES, INC (Attachments: # <u>1</u> Exhibit Ex. 1 - 2023.11.14 Metzker Dep. mini, # <u>2</u> Exhibit Ex. 2 - 2023.11.09 Hearing Transcript, # <u>3</u> Exhibit Ex. 3 - 2023.11.27 Hearing)(MORROW, JOHN) (Entered: 12/01/2023)					
12/01/2023	163	Proposed Findings of Fact and Conclusions of Law by NATERA, INC (Attachments: # Lexhibit 1. Metzker Depo. Tr., # 2 Exhibit 2. Stoll Depo. Tr.)(VAN ARNAM, ROBERT) (Entered: 12/01/2023)					
12/01/2023	164	Suggestion of Subsequently Decided Authority re 157 Suggestion of Subsequently Decided Authority Supplement with Redacted Order by Plaintiff NATERA, INC Attachments: #1 Exhibit 1. Redacted Injunction Order)(SHORES, ANDREW) (Entered: 2/01/2023)					
12/08/2023	165	RESPONSE re 164 Suggestion of Subsequently Decided Authority filed by NEOGENOMICS LABORATORIES, INC Replies due by 12/22/2023 (Attachments: # 1 Exhibit 1 - 2023-11-15 [DI675] Defs' Ntc of Sub Develop re Natera's PI Motion - Redacted)(MORROW, JOHN) (Entered: 12/08/2023)					
12/11/2023	166	NOTICE NOTICE OF FULL WITHDRAWAL OF NATERA'S MOTION TO SEAL (DOC. 132) AND PARTIAL WITHDRAWAL OF NATERA'S MOTION TO SEAL (DOC. 142) by NATERA, INC. re 160 Response in Support of Motion, 132 MOTION to Seal Unredacted Memorandum in Support of Motion to Dismiss Fifth Counterclaim and Strike Affirmative Defenses 10 & 11 [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).], 142 MOTION to Seal Reply in Support of its Motion for Preliminary Injunction Exhibits 1-8 attached to Nateras Reply, portions of the Reply Declaration of Dr. Michael L. Metzker, Ph.D, and Exhibit 1 attached to the Metzker Declaration. (SHORES, ANDREW) Modified on 12/12/2023 to add text to title of document. (at). (Entered: 12/11/2023)					
12/12/2023	167	WITHDRAWAL of Motion by Defendant NEOGENOMICS LABORATORIES, INC. re 127 MOTION to Seal <i>Second Amended Answer, Affirmative Defenses &amp; Counterclaims</i> [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] filed by NEOGENOMICS LABORATORIES, INC. (MORROW, JOHN) (Entered: 12/12/2023)					
12/13/2023	168	RESPONSE in Support re 142 MOTION to Seal Reply in Support of its Motion for Preliminary Injunction Exhibits 1-8 attached to Nateras Reply, portions of the Reply Declaration of Dr. Michael L. Metzker, Ph.D, and Exhibit 1 attached to the Metzker Declaration 144 SEALED Unredacted Documents (LR 5.4), filed by NEOGENOMICS LABORATORIES, INC Replies due by 12/27/2023 (Attachments: # 1 Affidavit Declaration of Alicia Olivo)(MORROW, JOHN) (Entered: 12/13/2023)					
12/27/2023	169	<b>ORDER</b> signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 12/27/2023; that Natera's preliminary injunction motion, Doc. <u>5</u> , is GRANTED. The preliminary injunction will issue separately. (sh) (Entered: 12/27/2023)					
12/27/2023	170	<b>ORDER</b> signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 12/27/2023, that the defendant's motion for sanctions, entitled "Motion to Bind Natera to its Asserted Conception Date," Doc. <u>137</u> , is DENIED. (lg) (Entered: 12/27/2023)					

/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 🗫 🕫: 283 Filed: 03/18/2024					
12/27/2023	171	PRELIMINARY INJUNCTION signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 12/27/2023; as set out herein. (sh) (Entered: 12/27/2023)					
12/27/2023	172	<b>ORDER</b> signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 12/27/2023, that the defendant's motion to dismiss, Doc. 51, is DENIED without prejudice to a motion for summary judgment. (lg) (Entered: 12/27/2023)					
12/27/2023	173	NOTICE OF APPEAL as to <u>171</u> Preliminary Injunction by NEOGENOMICS LABORATORIES, INC Filing fee \$ 605, receipt number BNCMDC-3729252. (REINES, EDWARD) (Entered: 12/27/2023)					
12/29/2023	<u>176</u>	MOTION to Stay <i>Preliminary Injunction</i> by NEOGENOMICS LABORATORIES, INC Response to Motion due by 1/19/2024 (Attachments: # 1 Text of Proposed Order) (REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	177	MEMORANDUM filed by Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. REDGENOMICS LABORATORIES, INC. re 176 MOTION to Stay <i>Preliminary Injunction</i> filed by NEOGENOMICS LABORATORIES, INC (REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	<u>178</u>	MOTION to Modify <u>171</u> Preliminary Injunction by NEOGENOMICS LABORATORIES, INC Response to Motion due by 1/19/2024 (Attachments: # <u>1</u> Text of Proposed Order)(REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	179	MEMORANDUM filed by Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. REOGENOMICS LABORATORIES, INC. re 178 MOTION to Modify 171 Preliminary Injunction filed by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Declaration of Sikri in Support)(REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	180	MOTION to Seal <i>Memorandums In Support of Motion to Stay and Motion to Modify</i> [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NEOGENOMICS LABORATORIES, INC Response to Motion due by 1/12/2024 (Attachments: # 1 Text of Proposed Order)(REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	181	MEMORANDUM filed by Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. re 180 MOTION to Seal <i>Memorandums In Support of Motion to Stay and Motion to Modify</i> [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] filed by NEOGENOMICS LABORATORIES, INC (Attachments: # 1 Declaration of Olivo in Support) (REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	182	SEALED UNREDACTED DOCUMENTS filed by Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. re 176 Motion to Stay. (REINES, EDWARD) (Entered: 12/29/2023)					
12/29/2023	183	SEALED UNREDACTED DOCUMENTS filed by Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC.					
		NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, NEOGEN					

:/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Page: 284 Filed: 03/18/2024 re 181 Memorandum. (Attachments: # 1 Declaration of Sikri in Support) (REINES, EDWARD) (Entered: 12/29/2023)
01/01/2024	184	NOTICE by NEOGENOMICS LABORATORIES, INC. re <u>177</u> Memorandum, <u>179</u> Memorandum, <u>182</u> SEALED Unredacted Documents (LR 5.4), <u>183</u> SEALED Unredacted Documents (LR 5.4), <i>of Corrected Citations</i> (REINES, EDWARD) (Entered: 01/01/2024)
01/02/2024		<b>TEXT ORDER</b> issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 1/2/24. No later than 9 am January 4, 2024, NeoGenomics SHALL file a supplement to the Motion to Modify, Doc. 178, in the form of proposed modifications to the language to Paragraph 3 of the preliminary injunction, that solves the clarity problems it identifies in the brief in support of the Motion to Modify, Doc. 179 at 8-10, (ECF pagination) and shall submit a copy in Word format to the case manager. No additional briefing, commentary, or explanation is authorized. Natera SHALL respond to the Motion to Modify, Doc. 178, and the proposed modification language, as quickly as possible so that the Court may promptly address the amount of the bond as required by Rule 65(c) and in any event no later than 5 pm January 5, 2024. Any reply shall be filed at the corresponding time on the next business day, but no later than 9 am January 8, 2024. (rw) Modified on 1/2/2024 to correct Doc. No. to Doc. 178 (rw). (Entered: 01/02/2024)
01/03/2024		<b>TEXT ORDER</b> in light of the proposed order at Doc. [178-1], Neogenomics does not have to file the supplement and Natera shall direct its response brief to the language proposed in Doc. [178-1] issued by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 1/3/24. (rw) (Entered: 01/03/2024)
01/04/2024	185	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 01/04/2024, that the motion to modify, Doc. 178, is GRANTED to this limited extent: The Preliminary Injunction, Doc. 171 at 2 4, is not in effect and its requirements are held in abeyance, pending a decision on security under Rule 65(c) and the posting of any security required. The motion to modify, Doc. 178, otherwise remains under advisement. The defendant's request for expedited briefing on the motion to stay is GRANTED. Natera shall file its response no later than January 8, 2024, and NeoGenomics may file a reply brief no later than January 11, 2024. The motion for stay, Doc. 176, otherwise remains under advisement. (at) (Entered: 01/04/2024)
01/04/2024		Set/Response Deadline re <u>176</u> MOTION to Stay <i>Preliminary Injunction</i> : Response to Motion due by 1/8/2024. Replies due by 1/11/2024. (at) (Entered: 01/04/2024)
01/04/2024	186	RESPONSE in Opposition re <u>178</u> MOTION to Modify <u>171</u> Preliminary Injunction filed by NEOGENOMICS LABORATORIES, INC. <i>REDACTED</i> filed by NATERA, INC Replies due by 1/5/2024 (Attachments: # <u>1</u> Exhibit 1. TD Cowen Estimate Changes, # <u>2</u> Exhibit 2. BTIG article re PI)(VAN ARNAM, ROBERT) (Entered: 01/04/2024)
01/04/2024	187	MOTION to Seal <i>Natera's Redacted Opp. to Motion to Modify</i> [If the party filing this motion is not the party claiming confidentiality, the party claiming confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NATERA, INC Response to Motion due by 1/18/2024 (Attachments: # 1 Text of Proposed Order)(VAN ARNAM, ROBERT) (Entered: 01/04/2024)
01/04/2024	188	SEALED UNREDACTED DOCUMENTS Natera's Opp. to Motion to Modify (highlighted) filed by Counter Defendants NATERA, INC., NATERA, INC., NATERA, INC., Plaintiff NATERA, INC., Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC. re 187 Motion to Seal, 186 Response in Opposition to Motion,. (Attachments: # 1 Exhibit 1. TD Cowen Estimate Changes, # 2 Exhibit 2. BTIG article re PI)(VAN ARNAM, ROBERT) (Entered: 01/04/2024)

7/24, 12:10 PM	Case	e: 24-1324	Document: 42-1	Prage: 285	Filed: 03/18/2024
01/05/2024	<u>189</u>	178 MOTION		ninary Injunctio	SORATORIES, INC., to Response to in filed by NEOGENOMICS intered: 01/05/2024)
01/08/2024		Opposition to accompanying James Malack Preliminary In Declaration of Declaration, <u>1</u> Memorandum Seal Natera's I	Nateras Motion for F 5 Declarations of Vish owski, 142 MOTION junction Exhibits 1-8 Dr. Michael L. Metz 78 MOTION to Mod 5 In Support of Motio	reliminary Injurnal Sikri, with at I to Seal Reply is attached to Natuker, Ph.D, and I I I I I I I I I I I I I I I I I I I	ntial portions of NeoGenomicss action (Opposition), as well as the tached Exhibits 7, 18, and 32, and in Support of its Motion for eras Reply, portions of the Reply Exhibit 1 attached to the Metzker tary Injunction, 180 MOTION to Sea Iotion to Modify and 187 MOTION to CHIEF DISTRICT JUDGE 024)
01/08/2024		NEOGENOM	ICS LABORATORII	ES, INC. <i>REDA</i>	Preliminary Injunction filed by CTED filed by NATERA, INC T) (Entered: 01/08/2024)
01/08/2024	191	not the party c response withi NATERA, INC	laiming confidentialin 14 days that includ	ty, the party clai es the materials on due by 1/22/2	190) [If the party filing this motion is ming confidentiality must file a required by L.R. 5.4(c)(3).] by 2024 (Attachments: # 1 Text of ed: 01/08/2024)
01/08/2024	192	by Counter De NATERA, INC Claimants NEC LABORATOR	efendants NATERA, E., Defendant NEOG OGENOMICS LABORIES, INC., NEOGE	INC., NATERA ENOMICS LAI DRATORIES, IN NOMICS LABO	cted Opp. to Motion to Stay PI filed, INC., NATERA, INC., Plaintiff BORATORIES, INC., Counter NC., NEOGENOMICS ORATORIES, INC. re 190 Response ARNAM, ROBERT) (Entered:
01/10/2024	193	01/10/2024, the subject to future this security, the preliminary patient blood is those same part NeoGenomics signed contract advisement. The immediately in late winter or a conference. In	at Security for the present at Security for the present and including injunction as follows amples taken before tients as authorized umay also offer RaDats. c. Otherwise, the he motion to stay, Deneet and confer about spring of 2025, and the	eliminary injunction and addition, Doc. 171, vs.: a. NeoGenon security is postered and a for use in the motion to modifice. 176, remains a proposed scheme Court will someeting and confidence.	THERINE C. EAGLES on etion, Doc. 171, is set at \$10,000,000 itional evidence. Once Natera posts will go into effect. The Court clarificatics may perform RaDaR tests on ed and on samples taken thereafter for preliminary injunction, Doc. 171. b. three trials for which it already has by, Doc. 178, remains under advisement. The parties should eduling order that will allow for trial on schedule an initial pretrial ferring the parties may immediately 1/2024)
01/11/2024	194	Claimants NECLABORATOR MOTION to S	OGENOMICS LABORIES, INC., NEOGE	ORATORIES, IN NOMICS LABO <i>nction</i> filed by N	ORATORIES, INC., Counter NC., NEOGENOMICS ORATORIES, INC., to Response to 1 NEOGENOMICS LABORATORIES
01/11/2024	<u>195</u>			1 "	o Its Motion to Stay PI [If the party stiality, the party claiming

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 @@e: 286 Filed: 03/18/2024
		confidentiality must file a response within 14 days that includes the materials required by L.R. 5.4(c)(3).] by NEOGENOMICS LABORATORIES, INC Response to Motion due by 1/25/2024 (Attachments: # 1 Text of Proposed Order)(REINES, EDWARD) (Entered: 01/11/2024)
01/11/2024	196	SEALED REPLY, filed by Defendant NEOGENOMICS LABORATORIES, INC., Counter Claimants NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., NEOGENOMICS LABORATORIES, INC., to Response to 176 MOTION to Stay <i>Preliminary Injunction</i> filed by NEOGENOMICS LABORATORIES, INC (REINES, EDWARD) (Entered: 01/11/2024)
01/12/2024		Motions Submitted: <u>176</u> MOTION to Stay Preliminary Injunction, <u>180</u> MOTION to Seal Memorandums In Support, <u>191</u> MOTION to Seal Opp. to Motion to Stay PI and <u>195</u> MOTION to Seal NeoGenomics's Redacted Reply to Its Motion to Stay PI to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES. (rw) (Entered: 01/12/2024)
01/12/2024	197	Preliminary Injunction BOND in the amount of \$10,000,000.00,Bond No. 1001160971, posted by NATERA, INC., signed by Judge CHIEF DISTRICT JUDGE CATHERINE C. EAGLES (Attachments: # 1 Exhibit A. Preliminary Injunction Bond)(VAN ARNAM, ROBERT) (Entered: 01/12/2024)
01/16/2024	198	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 01/16/2024. Natera has withdrawn its motion to seal certain briefs and exhibits, withdrawing its claim that information about its interactions with Dr. Mariam Jamal-Hanjani is confidential. See Doc. 166. Consistent with that withdrawal and without objection, the Clerk SHALL unseal the following: Doc. 134, Doc. 144-8, Doc. 149, Doc. 153-1, Doc. 153-9. NeoGenomics also withdraws it motion to seal at Doc. 127, which sought to seal parts of NeoGenomics Second Amended Answer and Counterclaims containing allegations about Natera's interactions with Dr. Jamal-Hanjani. See Doc. 167. To that end, and without objection, the Clerk SHALL unseal Doc. 128. (at) (Entered: 01/16/2024)
01/17/2024	199	NOTICE of Special Appearance by attorney MATTHEW I KREEGER on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3742815.) (KREEGER, MATTHEW) (Entered: 01/17/2024)
01/17/2024	200	NOTICE of Special Appearance by attorney DARALYN J. DURIE on behalf of Defendant NEOGENOMICS LABORATORIES, INC. (Filing fee \$ 25 receipt number ANCMDC-3742817.) (DURIE, DARALYN) (Entered: 01/17/2024)
01/26/2024	201	Subsequent NOTICE OF APPEAL as to <u>171</u> Preliminary Injunction, <u>169</u> Order on Motion for Preliminary Injunction, <u>193</u> Order, <u>172</u> Order on Motion to Dismiss by NEOGENOMICS LABORATORIES, INC Filing fee \$ 605, receipt number ANCMDC-3750096. Appeal Record due by 2/2/2024. (DURIE, DARALYN) (Entered: 01/26/2024)
01/29/2024	203	JOINT MOTION for Entry of Scheduling Order by NATERA, INC. (Attachments: # 1 Text of Proposed Order)(SHORES, ANDREW) (Entered: 01/29/2024)
01/30/2024		Motions Submitted: 203 JOINT MOTION <i>Ifor Entry of Scheduling Order</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (rw) (Entered: 01/30/2024)
01/31/2024	204	DEFENDANT'S PROOF OF SERVICE OF NOTICE OF PRELIMINARY INJUNCTION by NEOGENOMICS LABORATORIES, INC. re 169 Order on Motion for Preliminary Injunction, 193 Order, [DEFENDANT'S PROOF OF SERVICE OF NOTICE OF PRELIMINARY INJUNCTION] (Attachments: # 1 Exhibit A - Notice of Preliminary Injunction to Its Employees, # 2 Exhibit B - Notice of Preliminary Injunction to its Clinical Sales Team, # 3 Exhibit C - Notice of Preliminary Injunction to its External Clients, # 4 Exhibit D - Notice of Preliminary Injunction to its Employees Clarifying the

2/7/24, 12:10 PM	Cas	e: 24-1324 Document: 42-1 Prege: 287 Filed: 03/18/2024   Scope of the Injunction)(MORROW, JOHN) Modified text of title of document on 2/1/2024 (at). (Entered: 01/31/2024)					
02/02/2024	205	SCA FOR THE FEDERAL CIRCUIT - NOTICE OF DOCKETING. 24-1409 re 73 Notice of Appeal filed by NEOGENOMICS LABORATORIES, INC., 201 ubsequent Notice of Appeal, filed by NEOGENOMICS LABORATORIES, INC. (at) Entered: 02/02/2024)					
02/02/2024	206	ORDER signed by CHIEF DISTRICT JUDGE CATHERINE C. EAGLES on 02/02/2024, that the Court clarifies the preliminary injunction as follows: a. NeoGenomics may offer RaDaR for use in the clinical trial identified in Doc. 179 -1 at 4 under the contract that was finalized and awaiting signature at the time the motion for preliminary injunction was granted, once that contract is signed. b. NeoGenomics may offer RaDaR for use in the two clinical trials identified in Doc. 179 -1 at 4 that were approved by the ethics and review boards of the sponsoring institutions and that had finalized protocols in place when the motion for preliminary injunction was granted, upon the signing of contracts consistent with the terms approved and finalized at the time the motion for preliminary injunction was granted. c. NeoGenomics may not offer RaDaR for use in the fourth clinical trial identified in Doc. 179 -1 at 4 as it is still in the design phase. The motion to modify, Doc. 178, remains under advisement as to NeoGenomics' overbreadth contentions. (at) (Entered: 02/02/2024)					
02/05/2024	207	MOTION to Withdraw 125 Notice of Special Appearance <i>Joshua Harris</i> by NATERA, INC (Attachments: # 1 Text of Proposed Order)(SHORES, ANDREW) (Entered: 02/05/2024)					
02/06/2024		Motions Submitted: 207 MOTION to Withdraw 125 Notice of Special Appearance <i>Joshua Harris</i> to CHIEF DISTRICT JUDGE CATHERINE C. EAGLES- (rw) (Entered: 02/06/2024)					

PACER Service Center					
	Transaction Receipt				
	02/07/2024 10:27:37				
PACER Login:	mf000016	Client Code:			
Description:	Docket Report	Search Criteria:	1:23-ev-00629-CCE- JLW		
Billable Pages:	30	Cost:	3.00		

# IN THE UNITED STATES DISTRICT COURT FOR THE MIDDLE DISTRICT OF NORTH CAROLINA

NATERA, INC.,	)
Plaintiff,	)
v.	) C.A. No. 1:23-cv-629
NEOGENOMICS LABORATORIES, INC,	) ) JURY TRIAL DEMANDED )
Defendant.	)

# **COMPLAINT FOR PATENT INFRINGEMENT**

Natera, Inc. ("Natera") submits this Complaint against Defendant NeoGenomics Laboratories, Inc. ("NeoGenomics Laboratories"), and alleges as follows:

# **OVERVIEW OF THE ACTION**

1. This action arises under the patent laws of the United States, 35 U.S.C. §§ 1, et seq., from Defendant's infringement of Natera's U.S. Patent Nos. 11,530,454 (the "'454 Patent") and 11,519,035 (the "'035 Patent") (collectively, "Asserted Patents").

# **THE PARTIES**

- 2. Plaintiff Natera is a corporation organized and existing under the laws of the state of Delaware.
- 3. Founded in 2004, Natera (f.k.a. Gene Security Network) is a pioneering genetics and bioinformatics company with industry-leading healthcare products. Natera is dedicated to improving disease management for oncology, reproductive health, and organ transplantation. For well over a decade, Natera has been researching and developing non-

invasive methods for analyzing DNA in order to help patients and doctors manage diseases.

These ongoing efforts have given rise to a number of novel and proprietary genetic testing

services to assist with life-saving health management.

4. Natera's pioneering and ongoing innovation is especially evident in the area

of cell-free DNA ("cfDNA")-based testing. In the cfDNA field, Natera has developed

unique and highly optimized cfDNA-based processes that can be used to test non-

invasively for a range of conditions. Natera created an industry-leading cfDNA test,

Panorama, which showcases Natera's mastery of cfDNA in the field of non-invasive

prenatal testing. Natera has also applied its cfDNA testing platform to the challenge of

assessing cancer. Natera has developed its cfDNA technology for approval in the clinical

setting in order to provide patients with tools for early, clinically meaningful cancer

assessment. Natera was awarded approval for coverage by Medicare for multiple

indications.

5. In detecting and monitoring cancer, the use of non-invasive, blood-based

tests offers significant advantages over older methods, such as invasive tumor biopsy. But

the significant technological challenge is that such blood-based testing requires the

measurement of very small amounts of relevant genetic material—circulating-tumor DNA

("ctDNA")—within a much larger blood sample. Natera's approach combines proprietary

molecular biology and computational techniques to measure genomic variations in tiny

amounts of DNA, representing a fundamental advance in molecular biology.

Natera's cfDNA platform is the product of well over a decade of hard work

and investment of, on average, more than fifty million dollars per year in research and

development. Natera has expended substantial resources researching and developing its

technologies and establishing its reputation among physicians, insurers, and regulators as

a company committed to sound science and consistently accurate, reliable results. This

research, and the technological innovations resulting therefrom, are protected by a

substantial patent portfolio, with over 330 patents issued or pending worldwide, including

greater than 60 in the field of oncology.

7. Among these patented inventions include the '454 Patent and the '035 Patent,

each of which Defendant infringes. Defendant has used Natera's patented cfDNA

technology without permission and in violation of the patent laws.

8. Defendant NeoGenomics Laboratories, Inc. is a corporation organized and

existing under the laws of the State of Florida. Upon information and belief, NeoGenomics

Laboratories, Inc. operates laboratories throughout the United States, including in Durham,

North Carolina.

6.

9. Upon information and belief, Defendant NeoGenomics Laboratories, Inc. is

a corporate affiliate of both Inivata, Inc. and Inivata Ltd. (collectively, "Inivata").

Defendant NeoGenomics Laboratories, Inivata Inc., and Inivata Ltd. are each wholly-

owned subsidiaries of parent corporation NeoGenomics, Inc.

10. Defendant operates under and identifies with the trade name "NeoGenomics"

and "NeoGenomics Laboratories." Upon information and belief, Defendant directly or

indirectly makes, uses, offers to sell and/or sells in the United States an assay that infringes

at least one valid claim of each of the Asserted Patents, including in the State of North

Carolina and in this District, and otherwise purposefully directs activities to the same.

11. Instead of developing its own science for its cancer detection and monitoring

products, Defendant has unlawfully used Natera's patented technology, including in

connection with the RaDaR<sup>TM</sup> Minimum Residual Disease Assay and any other products

or assays that use the same or similar technology (collectively, "RaDaR" or "Accused

Assay").

**JURISDICTION AND VENUE** 

12. This Court has subject matter jurisdiction over the matters asserted herein

under 28 U.S.C. §§ 1331 and 1338(a).

13. NeoGenomics Laboratories is subject to this Court's personal jurisdiction at

least because it directs the operations of a CAP/CLIA-certified laboratory located in

Durham, North Carolina. In addition, NeoGenomics Laboratories is subject to this Court's

personal jurisdiction because, on information and belief, Defendant, directly or indirectly,

designs, develops, makes, uses, offers for sale, and/or sells the Accused Assay throughout

the United States and within this District. Defendant has infringed and continues to

infringe Natera's patents in this District by, among other things, engaging in infringing

conduct within and directed at or from this District and purposely and voluntarily placing

its infringing assay into the stream of commerce with the expectation that the infringing

assay will be used in this District.

14. Upon information and belief, venue is proper in this District pursuant to

28 U.S.C. §§ 1391 and 1400(b) because, among other things, NeoGenomics Laboratories

has a regular and established place of business in this District at its CAP/CLIA-certified

laboratory located in Durham, North Carolina and has committed acts of infringement at

this same location.

**BACKGROUND** 

15. Since 2004, Natera has been a global leader in genetic testing, including

cfDNA testing. Natera's mission is to improve the management of disease worldwide, and

it focuses on reproductive health, oncology, and organ transplantation. To improve the

management of these conditions, Natera has developed novel technologies to make

significant and accurate clinical assessments from the miniscule amounts of cfDNA present

in a single blood sample. These technologies include methods to manipulate cfDNA in

nonconventional ways in order to capture information about genetic variations in cfDNA

and usefully transform that information for noninvasive testing. Natera develops and

commercializes innovative, non-traditional methods for manipulating and analyzing

cfDNA, and offers a host of proprietary cfDNA genetic testing services to the public to

assist patients and doctors in evaluating and tracking critical health concerns.

16. Since its founding, Natera has researched, developed, and released ten

molecular tests with applications in prenatal, cancer, and organ transplants, many of which

are available through major health plans, or covered by Medicare or Medicaid, and

therefore available to most patients in need of those tests. Natera's tests have helped more

than four million individuals to date. Natera's robust laboratory now processes over

130,000 tests per month in the United States and internationally, improving the ability of

physicians to monitor and manage crucial health issues and patients to prosper around the

world.

17. Building on these innovations, in 2017, Natera launched its cfDNA test to

detect and monitor cancer, called Signatera®. Signatera is a personalized ctDNA

surveillance tool that detects molecular residual disease ("MRD") when assessing disease

recurrence or treatment response in solid tumors. Signatera is designed to screen for

multiple tumor-derived targets with each assay. It is optimized to detect extremely low

quantities of ctDNA and provides early knowledge of disease recurrence with a >99.5%

clinical test specificity.

18. MRD assessment has become a standard of care in the management of

patients with hematological malignancies, but until recently it has not been possible in solid

cancers due to technical limitations. Accurate MRD testing and molecular monitoring

offers the potential for physicians to change or escalate treatment in patients who are MRD-

positive, and to de-escalate or avoid unnecessary treatment in patients who are MRD-

negative. It also holds potential as a surrogate endpoint in clinical trials.

19. Natera's technology has been validated in multiple clinical studies. In

Cancer Research UK/University College London's Tracking Cancer Evolution through

Therapy ("TRACERx"), Natera's technology was used for the multi-year monitoring of

patient-specific single-nucleotide variants (SNVs) in plasma, to understand the evolution

of cancer mutations over time, and to monitor patients for disease recurrence. Results from

the first 100 early-stage lung cancer patients analyzed as part of the study were featured on

the cover of the May 2017 issue of Nature and showed that an early prototype version of

Signatera identified 43% more ctDNA-positive early-stage lung cancer cases than a generic

lung cancer panel and demonstrated its potential to detect residual disease, measure

treatment response, and identify recurrence up to 11 months earlier than the standard of

care, with a sensitivity of 93% at time of relapse.

20. The U.S. Food and Drug Administration ("FDA") recognized the importance

of Natera's Signatera and has granted it three "Breakthrough Device" designations

("BDDs") for multiple cancer types. The first BDD was awarded on May 6, 2019 to help

accelerate FDA assessment and review of Signatera as an in vitro companion diagnostic to

a certain cancer therapy. In 2021, the FDA granted two additional BDDs covering new

intended uses of the Signatera test to support its development through Phase III clinical

trials as a companion diagnostic to different cancer therapies.

21. Natera not only developed the pivotal technology for personalized MRD

testing, but has also invested considerable time and resources into developing the

personalized cancer monitoring market. It has lobbied and convinced physicians,

researchers, regulatory authorities and private payors on the feasibility of this new

personalized cancer monitoring technology through extensive studies, which led Medicare

to issue a draft Local Coverage Determination ("LCD") for Signatera in March 2019. In

its draft LCD, Medicare determined that "[t]he analytical validity and clinical validity of

minimal residual disease testing using cell-free DNA, and Signatera in particular, appears

to be well established based on available information for the test."

22. In addition, Signatera® achieved its first commercial policy coverage by Blue

Cross and Blue Shield of Louisiana, effective January 1, 2023, which covers Signatera

testing for plan members diagnosed with colorectal and muscle invasive bladder cancer

and for pan-cancer immunotherapy monitoring. Additionally, effective March 1, 2023,

Blue Shield of California, allows tumor-informed ctDNA testing with Signatera for

patients with stage I-IV cancer to provide information for performing targeted therapy

and/or monitoring for relapse or progression.

23. The clinical significance of Signatera has been acknowledged in over 40

peer-reviewed publications, including validation across multiple cancer types to detect

recurrence earlier compared to standard diagnostic tools.<sup>1</sup> At the American Society of

Clinical Oncology (ASCO) Annual Meeting held from June 2–6, 2023, Natera announced

its new data on Signatera across a wide variety of cancers, including colorectal (CRC),

lung, bladder, esophageal, pancreatic, melanoma, sarcoma and cholangiocarcinoma. The

results highlight the signficance of Signatera among the oncology commnity as a platform

to effectively use ctDNA to predict patient outcomes and assess treatment response for

both common and rare cancers.

<sup>1</sup> Reinert T, et al. Analysis of plasma cell-free DNA by ultradeep sequencing in patients with stages I to III colorectal cancer. JAMA Oncol. 2019;5(8):1124–1131; Coombes RC, et al. Personalized detection of circulating tumor DNA antedates breast cancer

metastatic recurrence. Clin Cancer Res. 2019;25(14):4255-4263.

24. Natera continues to be a global leader in cell-free DNA testing and was recognized among the top 10 most innovative health companies of 2022 for its development of Signatera.<sup>2</sup> Signatera was also a finalist of the Fierce Life Science Innovation Awards for Medical Device Innovation in 2021.<sup>3</sup> Signatera also won the 2021 MedTech Breakthrough Award for "Best New Technology Solution – Biopsy" in the medical device catetory.<sup>4</sup>

25. The Asserted Patents resulted from Natera's years-long research in developing innovative new methods for amplifying and sequencing cell-free DNA.

## **General Background of the Inventions**

### A. The '454 Patent

26. The '454 Patent, attached hereto as Exhibit 1, is entitled "Detecting Mutations and Ploidy in Chromosomal Segments" and issued from the USPTO on December 20, 2022. Natera owns the '454 Patent, including the right to enforce it and seek damages for infringement.

27. The '454 Patent claims methods for preparing plasma samples to detect ploidy of chromosome segments and single nucleotide variant ("SNV") mutations. The claimed methods perform whole genome sequencing of tumor samples and multiplex amplication of cfDNA isolated from plasma samples in order to detect SNV mutations. Independent claim 1 of the '454 Patent recites:

<sup>&</sup>lt;sup>2</sup> See Exhibit 27 at 2.

<sup>&</sup>lt;sup>3</sup> See Exhibit 28 at 2.

<sup>&</sup>lt;sup>4</sup> See Exhibit 29 at 5.

A method for preparing a plasma sample of a subject having cancer or suspected of having cancer or useful for detecting one or more single nucleotide variant (SNV) mutations in the plasma sample, the method comprising:

performing whole exome sequencing or whole genome sequencing on a tumor sample of the subject to identify a plurality of tumor-specific SNV mutations;

performing targeted multiplex amplification to amplify 10 to 500 target loci each encompassing a different tumor-specific SNV mutation from cell-free DNA isolated from a plasma sample of the subject or DNA derived therefrom to obtain amplicons having a length of 50-150 bases, wherein the target loci are amplified together in the same reaction volume; and

sequencing the amplicons to obtain sequence reads, and detecting one or more of the tumor-specific SNV mutations present in the cell-free DNA from the sequence reads, wherein the sequencing has a depth of read of at least 50,000 per target locus.

- 28. The claims of the '454 Patent are not directed to an abstract idea, natural law, or natural phenomenon. Rather, they are directed to an innovative method of sample preparation comprising both sequencing a tumor sample and amplifying nucleic acid samples obtained from blood plasma using synthetic primers and amplification products to provide a novel, innovative, and personalized solution to issues peculiar to the particular problem of detecting ploidy of chromosome segments and SNVs in patients with cancer. The claims of the '454 Patent cover methods of preparation of an unnatural preparation.
- 29. The '454 Patent claims are directed to specific, nonconventional, non-routine methods for overcoming previously unresolved problems in this area. For example, as of the date of the invention, it would not have been routine or conventional to perform the claimed techniques either individually or in combination, including: performing whole

exome sequencing or whole genome sequencing on a turmor sample to identify a plurality

of tumor-specific SNV mutations, performing targeted multiplex amplication to amplify

10 to 500 target loci each encompassing a different tumor-specific SNV mutation from

cfDNA to obtain amplicons having a length of 50-150 bases, wherein the target loci are

amplified together in the same reaction, and sequencing the amplicons to obtain sequence

reads, wherein the sequencing has a depth of read of at least 50,000 per target locus.

B. The '035 Patent

30. The '035 Patent, attached hereto as Exhibit 2, is entitled "Methods for

Simultaneous Amplification of Target Loci" and issued from the USPTO on December 6,

2022. Natera owns the '035 Patent, including the right to enforce it and seek damages for

infringement.

31. The '035 Patent claims methods for amplifying and sequencing multiple

nucleic acid regions of interest in one reaction volume. The claimed methods include

targeted amplification of a plurality of single nucleotide polymorphism (SNP) loci

associated with cancer on cell-free DNA in a single reaction volume and sequencing the

plurality of SNP loci by conducting massively parallel sequencing. Independent claim 1

of the '035 Patent recites:

A method for amplifying and sequencing DNA, comprising:

tagging isolated cell free DNA with one or more universal tail adaptors to generate tagged products, wherein the isolated cell-free DNA is isolated from a blood sample collected from a subject who is not a

pregnant women;

amplifying the tagged products one or more times to generate final amplification products, wherein one of the amplification steps comprises targeted amplification of a plurality of single nucleotide polymorphism (SNP) loci in a single reaction volume, wherein one of the amplifying steps introduces a barcode and one or more sequencing tags; and

sequencing the plurality of SNP loci on the cell free DNA by conducting massively parallel sequencing on the final amplification products, wherein the plurality of SNP loci comprises 25-2,000 loci associated with cancer.

- 32. The claims of the '035 Patent are not directed to an abstract idea, natural law, or natural phenomenon. Rather, they are directed to a novel and innovative laboratory method comprising tagging cfDNA with universal tail adaptors, amplifying the tagged cfDNA including targeted amplification of 25-2,000 SNP loci associated with cancer in a single reaction volume, and sequencing the plurality of SNP loci by conducting massively parallel sequencing.
- 33. The '035 Patent claims are directed to specific, nonconventional, non-routine methods for overcoming previously unresolved problems in this area. For example, as of the date of the invention, it would not have been routine or conventional to perform the claimed techniques either individually or in combination, including: tagging isolated cfDNA with one or more universal tail adaptors to generate tagged products; amplifying the tagged products one or more times to generate final amplification products, wherein one of the amplification steps comprises targeted amplification of a plurality of SNP loci in a single reaction volume, wherein one of the amplifying steps introduces a barcode and one or more sequencing tags; and sequencing the plurality of SNP loci on the cfDNA by

conducting massively parallel sequencing on the final amplification products, wherein the

plurality of SNP loci comprises 25-2,000 loci associated with cancer.

**DEFENDANT'S INFRINGING ACTS** 

34. The allegations provided below are exemplary and without prejudice to

Natera's infringement contentions. In providing these allegations, Natera does not convey

or imply any particular claim constructions or the precise scope of the claims. Natera's

claim construction contentions regarding the meaning and scope of the claim terms will be

provided under the Court's scheduling order and local rules.

35. The infringing products include, but are not limited to, RaDaR, and any other

infringing method, product, device, or test developed by Defendant that apply Natera's

patented methods for preparing, amplifying, sequencing, and analyzing cfDNA to detect

and monitor genes and genetic mutations associated with a patient's cancer.

36. As provided in more detail below, each element of at least one claim of

the '454 Patent is literally present in RaDaR or is literally practiced by the processes

through which RaDaR is practiced. To the extent that any element is not literally present

or practiced, each such element is present or practiced under the doctrine of equivalents.

37. As provided in more detail below, each element of at least one claim of

the '035 Patent is literally present in RaDaR or is literally practiced by the processes

through which RaDaR is practiced. To the extent that any element is not literally present

or practiced, each such element is present or practiced under the doctrine of equivalents.

38. Attached as Exhibits 3-4 are preliminary and exemplary claim charts

describing infringement of claim 1 of the '454 Patent and claim 1 of the '035 Patent,

respectively. Exhibits 5-16 are documents referenced in Exhibits 3 and 4, which

demonstrate examples of infringement. The claim charts are not intended to limit Natera's

right to modify the charts or allege that other products or activities of Defendant infringe

the identified claims or any other claims of the Asserted Patents or any other patents.

Defendant infringes more than one claim of the '454 Patent and infringes more than one

claim of the '035 Patent.

39. Exhibits 3-4 are hereby incorporated by reference in their entirety. Each

claim element in Exhibit 3 that is mapped to the Accused Assay shall be considered an

allegation within the meaning of the Federal Rules of Civil Procedure, and therefore a

response to each claim element is required. Each claim element in Exhibit 4 that is mapped

to the Accused Assay shall be considered an allegation within the meaning of the Federal

Rules of Civil Procedure, and therefore a response to each claim element is required.

40. Upon information and belief, NeoGenomics Laboratories designs, develops,

makes and uses, or directs or controls the design, development, make and use of, the

RaDaR assay by purposefully directing the activities at its CAP/CLIA-certified laboratory

in the Research Triangle Park (RTP) located at 8 Davis Drive, Durham, North Carolina.<sup>5</sup>

According to the CAP and CLIA certificates (attached hereto as Exhibits 18-19,

respectively), this laboratory is registered to Inivata, Inc., a corporate affiliate of

<sup>5</sup> See Exhibit 17 at 2.

NeoGenomics Laboratories, and the director of this laboratory is Dr. Siby Sebastian, Ph.D.<sup>6</sup>

On its website, NeoGenomics Laboratoris acknowledges that Dr. Siby Sebastian is also

"the Director of the NeoGenomics CAP/CLIA Clinical Laboratory at the Research Triangle

Park (RTP), North Carolina." Moreover, according to its 2022 Annual Report to

shareholders (attached hereto as Exhibit 21), NeoGenomics Laboratories stated that "[w]e

have a renewed focus on next-generation sequencing (NGS), minimal residual disease

(MRD) technology such as RaDaR . . . By centralizing the R&D function and integrating

Inivata, we believe we are now well-positioned to capitalize on the innovation of Inivata

while enhancing the development process for new product."8

41. Upon information and belief, NeoGenomics Laboratories performs, or

directs or controls the performance of, the RaDaR assay. In its published guide for patients,

NeoGenomics Laboratories states that RaDaR testing is "conducted on [a] blood sample in

our lab."9

42. Upon information and belief, NeoGenomics Laboratories also offers for sale

and sells, or directs or controls the offer for sale and sale of, the Accused Assay. On March

16, 2023, NeoGenomics began selling and offering to sell RaDaR in the United States.<sup>10</sup>

In its 2022 Annual Report to shareholders, NeoGenomics Laboratories stated that "[w]e

See Exhibits 18-19.

<sup>7</sup> See Exhibit 20.

<sup>8</sup> See Exhibit 21 at 7.

<sup>9</sup> See Exhibit 22 at 2.

<sup>10</sup> See Exhibit 23 at 1.

believe NeoGenomics is well-positioned to gain market share with our . . . RaDaR® assay

liquid biopsy test for MRD."11

43. Upon information and belief, NeoGenomics Laboratories induces Inivata to

perform any RaDaR tests that NeoGenomics is not performing itself. While RaDaR is

ordered through NeoGenomics Laboratories, its insurance reimbursement certifications

and approvals, including MolDX, list Inivata as authorized to perform RaDaR tests. 12 In

addition, the request forms for the RaDaR test report that RaDaR may be performed by

Inivata.<sup>13</sup>

44. NeoGenomics Laboratories is a direct competitor of Natera in the market for

personalized, tumor-informed recurrence monitoring specifically. In promoting its

infringing RaDaR assay, NeoGenomics Laboratories repeatedly undermined Signatera and

Natera to investors, including at the Goldman Sachs 44th Annual Global Healthcare

Conference held on June 13, 2023. NeoGenomics Laboratories touted its infringing and

competing RaDaR assay to investors as being "about 10x more sensitive than the other

product that's really commercially available in the market,"14 which is Signatera, even

though there is no clinical evidence or head-to-head study to substantiate this claim. On

information and belief, representatives of NeoGenomics Laboratories have repeated

<sup>11</sup> See Exhibit 21 at 9.

<sup>12</sup> See Exhibit 24.

<sup>13</sup> See Exhibit 25 at 2.

<sup>14</sup> See Exhibit 26 at 1.

similar misleading claims to potential customers, including current Signatera customers, in

promoting the infringing RaDaR assay.

45. On July 27, 2023, NeoGenomics Laboratories announced that RaDaR

obtained Medicare coverage for cancer recurrence detection and monitoring in patients

with breast cancer.<sup>15</sup>

46. NeoGenomics Laboratories has actual knowledge of the '454 Patent since at

least as early as December 20, 2022, when Natera filed an infringement lawsuit in the

District of Delware asserting the '454 Patent against two corporate affiliates of

NeoGenomics Laboratories: Inivata, Inc., and Inivata, Ltd. 16

47. NeoGenomics Laboratories has actual knowledge of the '035 Patent since at

least as early as the date of this Complaint.

48. NeoGenomics Laboratories has thus made extensive unauthorized use of

Natera's patented technology, including the technology described and claimed in the

Asserted Patents, causing Natera to suffer immediate and irreparable harm. Natera brings

this action to enforce its patents rights. Natera thus requests that this Court award it

damages sufficient to compensate for NeoGenomics Laboratories' infringement of the

Asserted Patents, find this case exceptional, award Natera its attorneys' fees and costs, and

<sup>15</sup> See Exhibit 30 at 1.

<sup>16</sup> See Natera, Inc. v. Inivata, Inc. et al, Case No. 1:22-cv-01609-UNA, Dkt. No. 1 (D.

Del. Dec. 20, 2022); see also Exhibit 12 at 115 (In its 2022 Annual Report,

NeoGenomics Laboratories acknowledged that Natera "filed a second patent

infringement complaint on December 20, 2022 against Inivata Limited and Inivata Inc.

alleging that RaDaR® minimal residual disease test infringes one patent.")

grant an injunction against NeoGenomics Laboratories to prevent any further infringement

of the Asserted Patents.

COUNT I: DIRECT INFRINGEMENT OF U.S. PATENT NO. 11,530,454

49. Natera incorporates by reference and re-alleges the foregoing paragraphs as

if fully set forth herein.

50. Natera is the owner of the '454 Patent, which was duly and legally issued by

the USPTO on December 20, 2022.

51. Defendant has infringed and continues to infringe at least one claim of

the '454 Patent pursuant to 35 U.S.C. § 271(a), literally or under the doctrine of equivalents,

by performing, or directing or controlling the performance of, the Accused Assay within

the United States and without authority.

52. Defendant's infringement has damaged and will continue to damage Natera,

which is entitled to recover the damages resulting from Defendant's wrongful acts in an

amount to be determined at trial, and in any event no less than a reasonable royalty.

53. Moreover, Defendant's infringement has caused, and will continue to cause,

irreparable injury to Natera, for which damages are an inadequate remedy, unless

Defendant, including its corporate affiliates and subsidiaries, is enjoined from any and all

activities that would infringe the claims of the '454 Patent.

COUNT II: DIRECT INFRINGEMENT OF U.S. PATENT NO. 11,519,035

54. Natera incorporates by reference and re-alleges the foregoing paragraphs as

if fully set forth herein.

55. Natera is the owner of the '035 Patent, which was duly and legally issued by

the USPTO on December 6, 2022.

56. Defendant has infringed and continues to infringe at least one claim of

the '035 Patent pursuant to 35 U.S.C. § 271(a), literally or under the doctrine of equivalents,

by performing, or directing or controlling the performance of, the Accused Assay within

the United States and without authority.

57. Defendant's infringement has damaged and will continue to damage Natera,

which is entitled to recover the damages resulting from Defendant's wrongful acts in an

amount to be determined at trial, and in any event no less than a reasonable royalty.

58. Moreover, Defendant's infringement has caused, and will continue to cause,

irreparable injury to Natera, for which damages are an inadequate remedy, unless

Defendant, including its corporate affiliates and subsidiaries, is enjoined from any and all

activities that would infringe the claims of the '035 Patent.

COUNT III: INDIRECT INFRINGEMENT OF U.S. PATENT NO. 11,530,454

59. Natera incorporates by reference and re-alleges the foregoing paragraphs as

if fully set forth herein.

60. NeoGenomics Laboratories has indirectly infringed and continues to

indirectly infringe at least one claim of the '454 Patent pursuant to 35 U.S.C. § 271(b),

literally or under the doctrine of equivalents, by inducing its corporate affiliate Inivata to

perform the Accused Assay within the United States without authority. NeoGenomics

Laboratories engaged in such inducement having knowledge of the '454 Patent at least as

of December 20, 2022 when Natera filed a lawsuit alleging infringement of the '454 Patent

by Inivata in the District of Delaware.<sup>17</sup> Furthermore, NeoGenomics Laboratories knew or

should have known that its actions of selling or offering to sell the Accused Assay would

induce direct infringement by Inivata and intended that its actions would induce direct

infringement by Inivata. NeoGenomics Laboratories offers for sale the Accused Assay via

its website specifically intending that Inivata perform the Accused Assay in NeoGenomics

Laboratories' CAP/CLIA-certified laboratory in Durham, NC upon purchase by a

customer.

61. As a direct and proximate result of NeoGenomics Laboratories' indirect

infringement by inducement of the '454 Patent, Natera has been and continues to be

damaged. Thus, Natera is entitled to recover the damages resulting from NeoGenomics

Laboratories' wrongful acts in an amount to be determined at trial, and in any event no less

than a reasonable royalty.

62. Moreover, Defendant's infringement has caused, and will continue to cause,

irreparable injury to Natera, for which damages are an inadequate remedy, unless

NeoGenomics Laboratories is enjoined from any and all activities that would infringe the

claims of the '454 Patent.

<sup>17</sup> Natera, Inc. v. Inivata, Inc. et al., Case No. 22-cv-1609-GBW, Dkt. No. 1 (D. Del.

Dec. 20, 2022).

COUNT IV: INDIRECT INFRINGEMENT OF U.S. PATENT NO. 11,519,035

63. Natera incorporates by reference and re-alleges the foregoing paragraphs as

if fully set forth herein.

64. NeoGenomics Laboratories has indirectly infringed and continues to

indirectly infringe at least one claim of the '035 Patent pursuant to 35 U.S.C. § 271(b),

literally or under the doctrine of equivalents, by inducing its corporate affiliate Inivata to

perform the Accused Assay within the United States without authority. NeoGenomics

Laboratories engaged in such inducement having knowledge of the '035 Patent, at least as

of the service of the present Complaint. Furthermore, NeoGenomics Laboratories knew or

should have known that its actions of selling or offering to sell the Accused Assay would

induce direct infringement by Inivata and intended that its actions would induce direct

infringement by Inivata. NeoGenomics Laboratories offers for sale the Accused Assay via

its website specifically intending that Inivata perform the Accused Assay in NeoGenomics

Laboratories' CAP/CLIA-certified laboratory in Durham, NC upon purchase by a

customer.

65. As a direct and proximate result of NeoGenomics Laboratories' indirect

infringement by inducement of the '035 Patent, Natera has been and continues to be

damaged. Thus, Natera is entitled to recover the damages resulting from NeoGenomics

Laboratories' wrongful acts in an amount to be determined at trial, and in any event no less

than a reasonable royalty.

21

Case 1:23-cv-00629-CCE-JLW Document 1 Filed 07/28/23 Page 21 of 26 Appx 305

66. Moreover, Defendant's infringement has caused, and will continue to cause, irreparable injury to Natera, for which damages are an inadequate remedy, unless

NeoGenomics Laboratories is enjoined from any and all activities that would infringe the

claims of the '035 Patent.

COUNT V: WILLFUL INFRINGEMENT OF U.S. PATENT NO. 11,530,454

67. Natera incorporates by reference and re-alleges the foregoing paragraphs as

if fully set forth herein.

68. NeoGenomics Laboratories has acted willfully and egregiously in

performing the acts of infringement identified in this Complaint. NeoGenomics

Laboratories' infringement of the '454 Patent has been and is deliberate and willful and

constitutes egregious misconduct. NeoGenomics Laboratories engaged in such conduct

having knowledge of the '454 Patent at least as of December 20, 2022 when Natera filed a

lawsuit alleging infringement of the '454 Patent by Inivata in the District of Delaware.<sup>18</sup>

Furthermore, NeoGenomics Laboratories knew or should have known that its actions of

selling or offering to sell the Accused Assay would induce direct infringement by Inivata

and intended that its actions would induce direct infringement by Inivata. NeoGenomics

Laboratories offers for sale the Accused Assay via its website specifically intending that

Inivata perform the Accused Assay in NeoGenomics Laboratories' CAP/CLIA-certified

laboratory in Durham, NC upon purchase by a customer. In performing the acts of

<sup>18</sup> Natera, Inc. v. Inivata, Inc. et al., Case No. 22-cv-1609-GBW, Dkt. No. 1 (D. Del.

Dec. 20, 2022).

infringement identified in this Complaint, NeoGenomics Laboratories has been willfully

blind to its ongoing infringement.

69. As a direct and proximate result of NeoGenomics Laboratories' willful

infringement, Natera has been and continues to be damaged. Thus, Natera is entitled to

recover the damages resulting from NeoGenomics Laboratories' wrongful acts in an

amount to be determined at trial. Natera deserves treble damages and the reimbursement

of its fees and costs as set forth in 35 U.S.C. §§ 284 and 285.

70. Moreover, Defendant's infringement has caused, and will continue to cause,

irreparable injury to Natera, for which damages are an inadequate remedy, unless

NeoGenomics Laboratories is enjoined from any and all activities that would infringe the

claims of the '454 Patent.

COUNT VI: WILLFUL INFRINGEMENT OF U.S. PATENT NO. 11,519,035

71. Natera incorporates by reference and re-alleges the foregoing paragraphs as

if fully set forth herein.

72. NeoGenomics Laboratories has acted willfully and egregiously in

performing the acts of infringement identified in this Complaint. NeoGenomics

Laboratories' infringement of the '035 Patent has been and is deliberate and willful and

constitutes egregious misconduct. NeoGenomics Laboratories engaged in such conduct

having knowledge of the '035 Patent, at least as of the service of the present Complaint.

Furthermore, NeoGenomics Laboratories knew or should have known that its actions of

selling or offering to sell the Accused Assay would induce direct infringement by Inivata

and intended that its actions would induce direct infringement by Inivata. NeoGenomics

Laboratories offers for sale the Accused Assay via its website specifically intending that

Inivata perform the Accused Assay in NeoGenomics Laboratories' CAP/CLIA-certified

laboratory in Durham, NC upon purchase by a customer. In performing the acts of

infringement identified in this Complaint, NeoGenomics Laboratories has been willfully

blind to its ongoing infringement.

73. As a direct and proximate result of NeoGenomics Laboratories' willful

infringement, Natera has been and continues to be damaged. Thus, Natera is entitled to

recover the damages resulting from NeoGenomics Laboratories' wrongful acts in an

amount to be determined at trial. Natera deserves treble damages and the reimbursement

of its fees and costs as set forth in 35 U.S.C. §§ 284 and 285.

74. Moreover, Defendant's infringement has caused, and will continue to cause,

irreparable injury to Natera, for which damages are an inadequate remedy, unless

NeoGenomics Laboratories is enjoined from any and all activities that would infringe the

claims of the '035 Patent.

PRAYER FOR RELIEF

WHEREFORE, Natera respectfully requests the following relief:

1. A judgment that Defendant has infringed the '454 Patent and the'035 Patent

literally or under the doctrine of equivalents;

2. An order enjoining Defendant and their respective officers, directors, agents,

servants, affiliates, employees, divisions, branches, subsidiaries, parents, and all others

acting on behalf of or in active concert or participation therewith, from further infringement

of the '454 Patent and the '035 Patent;

3. An award of damages sufficient to compensate Natera for Defendant's

infringement under 35 U.S.C. § 284;

4. A declaration that Defendant's infringement of '454 Patent and the'035

Patent has been willful and deliberate, and an increase to the award of damages of three

times the amount found or assessed by the Court, in accordance with 35 U.S.C. § 284;

5. A determination that this is an exceptional case under 35 U.S.C. § 285 and

that Natera be awarded attorneys' fees;

6. Costs and expenses in this action;

7. An award of prejudgment and post-judgment interest; and

8. Such other and further relief as the Court may deem just and proper.

**DEMAND FOR JURY TRIAL** 

Pursuant to Rule 38(b) of the Federal Rules of Civil Procedure, Natera respectfully

demands a trial by jury on all triable issues.

Dated: July 28, 2023

Resepctfully Submitted,

WILLIAMS MULLEN

By: <u>/s/Robert C. Van Arnam</u>

Robert C. Van Arnam (NC #28838)

Andrew R. Shores (NC #46600)

301 Fayetteville Street, Suite 1700

> Raleigh, NC 27601 Telephone: (919) 981-4015 Fax: (919) 981-4300

Email: rvanarnam@williamsmullen.com ashores@williamsmullen.com

Attorneys for Plaintiff Natera, Inc.

### OF COUNSEL:

Kevin P.B. Johnson Victoria F. Maroulis QUINN EMANUEL URQUHART & SULLIVAN, LLP 555 Twin Dolphin Shores, 5th Floor Redwood Shores, CA 94065 (650) 801-5000

Andrew M. Holmes
Sam S. Stake
QUINN EMANUEL URQUHART
& SULLIVAN, LLP
50 California Street, 22nd Floor
San Francisco, CA 94111
(415) 875-6600

Sandra Haberny, Ph.D.

QUINN EMANUEL URQUHART

& SULLIVAN, LLP

865 South Figueroa Street, 10th Floor
Los Angeles, CA 90017

(213) 443-3000

# **EXHIBIT 1**

Document: 42-1 Case: 24-1324 Page: 315 Filed: 03/18/2024



## (12) United States Patent Babiarz et al.

#### US 11,530,454 B2 (10) Patent No.: \*Dec. 20, 2022 (45) **Date of Patent:**

### (54) DETECTING MUTATIONS AND PLOIDY IN CHROMOSOMAL SEGMENTS

- (71) Applicant: Natera, Inc., San Carlos, CA (US)
- (72) Inventors: Joshua Babiarz, Castro Valley, CA (US); Tudor Pompiliu Constantin, Berkeley, CA (US); Lane A. Eubank, San Carlos, CA (US); George Gemelos, Portland, OR (US); Matthew Micah Hill, Belmont, CA (US); Huseyin Eser Kirkizlar, Los Angeles, CA (US); Matthew Rabinowitz, San Francisco, CA (US); Onur Sakarva, Redwood City, CA (US); Styrmir Sigurjonsson, San Jose, CA (US); Bernhard Zimmermann, Manteca, CA (US)
- (73) Assignee: Natera, Inc., San Carlos, CA (US)
- (\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer

(21) Appl. No.: 17/738,354

(22) Filed: May 6, 2022

#### (65)**Prior Publication Data**

US 2022/0282335 A1 Sep. 8, 2022

### Related U.S. Application Data

- (63) Continuation of application No. 17/692,469, filed on Mar. 11, 2022, which is a continuation of application No. 15/898,145, filed on Feb. 15, 2018, now Pat. No. 11,319,595, which is a continuation of application No. 14/692,703, filed on Apr. 21, 2015, now Pat. No. 10,179,937.
- (60) Provisional application No. 62/148,173, filed on Apr. 15, 2015, provisional application No. 62/147,377, filed on Apr. 14, 2015, provisional application No. 62/146,188, filed on Apr. 10, 2015, provisional application No. 62/066,514, filed on Oct. 21, 2014, provisional application No. 61/994,791, filed on May 16, 2014, provisional application No. 61/987,407, filed on May 1, 2014, provisional application No. 61/982,245, filed on Apr. 21, 2014.

(51)	Int. Cl.	
	C12Q 1/6886	(2018.01)
	G06N 20/00	(2019.01)
	G16B 15/00	(2019.01)
	G16B 25/00	(2019.01)
	G16B 40/00	(2019.01)
	C12Q 1/6869	(2018.01)
	G16B 40/20	(2019.01)
	G16B 20/10	(2019.01)
	G16B 20/00	(2019.01)
	G16B 20/20	(2019.01)
	G16Z 99/00	(2019.01)

G16H 50/20 (2018.01)G06N 7/00 (2006.01)G16H 10/40 (2018.01)G16B 25/20 (2019.01)

(52) U.S. Cl. CPC ...... C12Q 1/6886 (2013.01); C12Q 1/6869 (2013.01); G06N 7/005 (2013.01); G06N **20/00** (2019.01); **G16B** 15/00 (2019.02); G16B 20/00 (2019.02); G16B 20/10 (2019.02); G16B 20/20 (2019.02); G16B 20/20 (2019.02); G16B 25/00 (2019.02); G16B 40/00 (2019.02); G16B 40/20 (2019.02); G16H 10/40 (2018.01); G16H 50/20 (2018.01); G16Z 99/00 (2019.02); C12Q 2539/10 (2013.01); C12Q 2600/156 (2013.01); C12Q 2600/158 (2013.01); C12Q 2600/16 (2013.01); C12Q 2600/172 (2013.01); G16B 25/20 (2019.02)

(58) Field of Classification Search See application file for complete search history.

#### References Cited (56)

### U.S. PATENT DOCUMENTS

3,957,654 A	5/1976	Ayres
4,040,785 A	8/1977	Kim et al.
4,683,195 A	7/1987	Mullis et al.
4,942,124 A	7/1990	Church et al.
5,486,477 A	1/1996	Carver
5,635,366 A	6/1997	Cooke et al.
5,648,220 A	7/1997	Bianchi et al.
5,716,776 A	2/1998	Bogart
5,736,033 A	4/1998	Coleman et al.
5,753,467 A	5/1998	Jensen et al.
	(Con	tinued)

### FOREIGN PATENT DOCUMENTS

CN	1650032 A	8/2005
CN	1674028 A	9/2005
	(Cont	inued)

### OTHER PUBLICATIONS

US 8,501,409 B2, 08/2013, Simen et al. (withdrawn)

Forshew et al. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA Science Translational Medicine vol. 4 article 136ra68 and Supplementary Material (Year: 2012).\*

Benesova et al. Mutation-based detection and monitoring of cellfree tumor DNA in peripheral blood of cancer patients Analytical Biochemistry vol. 433, pp. 227-234 (Year: 2013).\*

(Continued)

Primary Examiner — John S Brusca

#### (57)ABSTRACT

The invention provides methods, systems, and computer readable medium for detecting ploidy of chromosome segments or entire chromosomes, for detecting single nucleotide variants and for detecting both ploidy of chromosome segments and single nucleotide variants. In some aspects, the invention provides methods, systems, and computer readable medium for detecting cancer or a chromosomal abnormality in a gestating fetus.

28 Claims, 105 Drawing Sheets

(56)		Referen	ces Cited		8,318,434			Cuppens et al. Andersen et al.
	U.S.	PATENT	DOCUMENTS		8,323,897 8,372,584	B2	2/2013	Shoemaker et al.
					8,389,557			Singh et al.
	,467 A		Mascarenhas		8,389,578			Went et al.
	,033 A	12/1998			8,450,063 8,467,976			Dube et al. Lo et al.
	917 A		Comanor et al. Gill et al.		8,515,679			Rabinowitz et al.
	,734 A ,170 A		Stroun et al.		8,532,930			Rabinowitz et al.
	,223 A		Whiteley et al.		8,679,741			Hoyal-Wrightson et al.
	,602 A		Hyland et al.		8,682,592			Rabinowitz et al.
	,790 A		Pinkel et al.		8,703,652			Quake et al.
	,148 A		Stewart et al.		8,706,422 8,748,103			Lo et al. Faham et al.
	,611 A	12/1999			8,822,153			Hayes et al.
	,128 A ,454 A		Veltri et al. Lipshutz et al.		8,825,412			Rabinowitz et al.
	0,029 A		Lapidus et al.		9,005,894			Ladner et al.
	,635 A		Herren et al.		9,051,602			Oliphant et al.
	,120 A		Lizardi		9,085,798		7/2015	
	,496 A		Brown et al.		9,206,475 9,228,234			Gerdes et al. Rabinowitz et al.
	5,504 A		Gocke et al.		9,228,234			Rava et al.
	,349 B1 ,472 B1		Ginzinger Landegren et al.		9,404,150			Lee et al.
	558 B1		Shuber et al.		9,424,392			Rabinowitz et al.
	,603 B1		Mahtani		9,453,257			Hoyal-Wrightson et al.
	,540 B1	7/2001	Lo et al.		9,476,095			Vogelstein et al.
	,077 B1		Shuber et al.		9,487,829 9,493,828			Vogelstein et al. Rava et al.
	,179 B1		Kopreski		9,493,828			Faham et al.
	,167 B1 ,706 B1		Pinkel et al.		9,598,731			Talasaz
	,700 B1 ,235 B1		Vogelstein et al. Schumm et al.		9,677,118			Zimmermann et al.
	,135 B1		Parrott et al.		10,011,870			Zimmermann et al.
	,451 B1		Marmaro et al.		10,017,810			Infrate et al.
	,137 B2		Dean et al.		10,041,127			Talasaz Rabinowitz et al.
	,140 B1		Hartley et al.		10,061,890 10,081,839			Rabinowitz et al.
	,140 B1 ,491 B2		Goldsborough Pavlovic et al.		10,083,273			Rabinowitz et al.
	,491 B2 3,412 B2		Willis et al.		10,174,369			Rabinowitz et al.
	,028 B2		Lo et al.		10,179,937			Babiarz et al.
	,487 B1	10/2005	Barany et al.		10,227,652			Rabinowitz et al.
	3,211 B2		Vingerhoets et al.		10,229,244 10,240,202		3/2019	Rabinowitz et al.
	,847 B1	11/2005			10,260,096			Rabinowitz et al.
	5,739 B2 5,517 B1		Schadt et al. Denton et al.		10,266,893			Rabinowitz et al.
	3,616 B1		Larder et al.		10,308,981		6/2019	Sparks et al.
	,663 B2		Godfrey et al.		10,316,362			Babiarz et al.
	,656 B2		Nolan et al.		10,351,906			Zimmermann et al.
	3,764 B2		Vaisberg et al.		10,385,396 10,392,664			Mitchell et al. Rabinowitz et al.
	,485 B2 ,277 B2		Bornarth et al. Dhallan		10,450,597			Iafrate et al.
	,764 B2		Gocke et al.		10,472,680			Mitchell et al.
	1,118 B1		Mullah et al.		10,522,242			Rabinowitz et al.
7,442	,506 B2	12/2008			10,526,658			Babiarz et al.
	,273 B2		Jones et al.		10,538,814 10,557,172			Babiarz et al. Babiarz et al.
	,576 B2 ,399 B2		Lo et al.		10,597,708			Zimmermann et al.
	,399 B2 ,325 B2		Cantor et al. Cantor et al.		10,597,709			Zimmermann et al.
	3,367 B2		Lo et al.		10,597,723		3/2020	Babiarz et al.
	,370 B2		Dhallan		10,655,180			Babiarz et al.
	,463 B2		Gormley et al.		10,711,309			Rabinowitz et al.
	,798 B2		Cantor et al.		10,731,220 10,774,380			Babiarz et al. Ryan et al.
	,720 B2 ,393 B2		Dhallan Lyamichev et al.		10,793,912			Babiarz et al.
	,418 B2	9/2010			10,894,976		1/2021	Stray et al.
	,282 B2	11/2010			11,111,543			Rabinowitz et al.
	,647 B2	11/2010	Hahn et al.		11,111,544			Rabinowitz et al.
	,609 B2		Rubin et al.		11,111,545 11,130,995		9/2021	Babiarz et al. Quake et al.
	3,017 B2	8/2011			11,130,993			Babiarz G16B 40/20
	3,018 B2 3,128 B2		Quake et al. Rabinowitz		11,371,100			Babiarz C12Q 1/6886
	,719 B2		Drmanac et al.	20	001/0051341			Lo et al.
	,912 B2		Kapur et al.	20	01/0053519	A1	12/2001	Fodor et al.
	,370 B2		Oeth et al.		002/0006622			Bradley et al.
	,389 B2		Shoemaker et al.		002/0107640			Ideker et al.
	5,503 B2		Faham et al.		002/0119478			Umansky et al.
	,415 B2 ,076 B2		Fan et al. Fan et al.		002/0182622			Nakamura et al. Markowitz et al.
	,076 B2 1,187 B2		Fan et al. Fernando		03/0009293			Langmore et al.
	3,430 B2		Chuu et al.		03/0044388			Lo et al.
5,510	,			2.				

(56)	Referen	ices Cited		007/0059700 007/0059707			Tao et al.
U.S	. PATENT	DOCUMENTS	2	007/0122805	A1	5/2007	Cantor et al.
2002/0065525	4/2002	77 1 · 1		007/0128624 007/0134658			Gormley et al. Bohmer et al.
2003/0065535 A1 2003/0077586 A1		Karlov et al. Pavlovic et al.		007/0178478			Dhallan
2003/0087276 A1		Kopreski et al.		007/0178501			Rabinowitz et al.
2003/0101000 A1		Bader et al.		007/0184467 007/0202525			Rabinowitz et al. Quake et al.
2003/0119004 A1 2003/0138780 A1		Wenz et al. Gill et al.		007/0202525			Yamanishi et al.
2003/0191005 A1		Coelho et al.		007/0207466			Cantor et al.
2003/0211489 A1		Shen et al.		007/0212689 007/0243549			Bianchi et al. Bischoff
2003/0228613 A1 2003/0232348 A1		Bornarth et al. Jones et al.	_	007/0259351		11/2007	
2003/0232353 A1		Kennedy et al.		008/0020390			Mitchell
2003/0235848 A1		Neville et al.		008/0026390 008/0038733			Stoughton et al. Bischoff et al.
2004/0009518 A1 2004/0033596 A1		Lo et al. Threadgill et al.		008/0050739			Stoughton et al.
2004/0067493 A1		Matsuzaki et al.		008/0070792		3/2008	Stoughton
2004/0096874 A1		Neville et al.		008/0071076 008/0085836			Hahn et al. Kearns et al.
2004/0115629 A1 2004/0117346 A1		Panzer et al. Stoffel et al.		008/0090239			Shoemaker et al.
2004/011/340 A1 2004/0126760 A1		Broude		008/0096766		4/2008	
2004/0137470 A1		Dhallan et al.		008/0102455 008/0138809		5/2008	Poetter Kapur et al.
2004/0146866 A1 2004/0157243 A1	7/2004 8/2004	Fu Huang et al.		008/0164204			Hatamian et al.
2004/0137243 A1 2004/0185495 A1		Schueler et al.	2	008/0182244	A1	7/2008	Tatas et al.
2004/0197797 A1	10/2004	Inoko et al.		008/0193927			Mann et al. Shoemaker et al.
2004/0209299 A1 2004/0229231 A1		Pinter et al. Frudakis et al.		008/0220422 008/0234142		9/2008	
2004/0229231 A1 2004/0236518 A1		Pavlovic et al.		008/0243398		10/2008	Rabinowitz et al.
2004/0259100 A1	12/2004	Gunderson et al.		008/0280292			Wangh et al.
2005/0009069 A1		Liu et al.		008/0286783 008/0299562			Hosono et al. Oeth et al.
2005/0043894 A1 2005/0049793 A1		Fernandez Paterlini-brechot		008/0305473		12/2008	Chowdary et al.
2005/0053950 A1	3/2005	Ubani et al.		009/0023190			Lao et al.
2005/0064476 A1		Huang et al.		009/0029377 009/0053719			Lo et al. Lo et al.
2005/0079521 A1 2005/0079535 A1		Beaulieu et al. Kirchgesser et al.		009/0087847			Lo et al.
2005/0123914 A1	6/2005	Katz et al.		009/0098534			Weier et al.
2005/0130173 A1		Leamon et al.		009/0099041 009/0143570			Church et al. Jiang et al.
2005/0142577 A1 2005/0144664 A1		Jones et al. Smith et al.		009/0176662		7/2009	Rigatti et al.
2005/0164241 A1		Hahn et al.		009/0221620			Luke et al.
2005/0164252 A1		Yeung		009/0228299 009/0263800			Kangarloo et al. Wohlgemuth et al.
2005/0216207 A1 2005/0221341 A1		Kermani Shimkets et al.		009/0280479			Hoon et al.
2005/0227263 A1	10/2005	Green et al.		009/0317817			Oeth et al.
2005/0250111 A1		Xie et al.		010/0012598 010/0035232			Dicesare et al. Ecker et al.
2005/0255508 A1 2005/0272073 A1		Casey et al. Vaisberg et al.		010/0041048		2/2010	Diehl et al.
2005/0282185 A1		Lo et al.		010/0086914			Bentley et al.
2006/0019278 A1		Lo et al.		010/0105049 010/0112575			Ehrich et al. Fan et al.
2006/0040300 A1 2006/0046258 A1		Dapprich et al. Lapidus et al.		010/0112586			Stoughton et al.
2006/0051799 A1		Iwaki et al.		010/0112590			Lo et al.
2006/0052945 A1		Rabinowitz et al.		010/0120038 010/0124751			Mir et al. Quake et al.
2006/0057618 A1 2006/0068369 A1		Piper et al. Coelho et al.	2	010/0129792	A1	5/2010	Makrigiorgos et al.
2006/0068394 A1	3/2006	Langmore et al.		010/0129874			Mitra et al.
2006/0088574 A1		Manning et al.		010/0138165 010/0155343			Fan et al. Battles et al.
2006/0094010 A1 2006/0099614 A1		Giles et al. Gill et al.	2	010/0171954	A1	7/2010	Quake et al.
2006/0121452 A1	6/2006	Dhallan		010/0173394			Colston et al.
2006/0134662 A1		Pratt et al. Sher et al.		010/0184043 010/0184069			Mitchell et al. Fernando et al.
2006/0141499 A1 2006/0229823 A1	8/2006			010/0184152			Sandler
2006/0210997 A1	9/2006	Myerson et al.		010/0196892			Quake et al.
2006/0216153 A1		Wobben et al.		010/0203538 010/0216151			Dube et al. Lapdus et al.
2006/0216738 A1 2006/0228721 A1		Wada et al. Leamon et al.		010/0216151			Lapidus et al.
2006/0248031 A1	11/2006	Kates et al.	2	010/0248231	A1	9/2010	Wei et al.
2006/0281105 A1		Li et al.		010/0255492			Quake et al.
2006/0292599 A1 2007/0020640 A1		Ritz et al. McCloskey et al.		010/0256013 010/0273159		10/2010 10/2010	Quake et al. Melo
2007/0020040 A1 2007/0027636 A1		Rabinowitz		010/02/3139			May et al.
2007/0031857 A1	2/2007	Makarov et al.	2	010/0273678	A1	10/2010	Alexandre et al.
2007/0037166 A1		Wohlgemuth et al.		010/0285478			Chen et al.
2007/0042384 A1	2/2007	Li et al.	2	010/0285537	AI	11/2010	Zimmermann

(56)	Referen	ices Cited		2013/0253369			Rabinowitz et al.
IIS	PATENT	DOCUMENTS		2013/0261004 2013/0274116		10/2013 10/2013	
0.0	. 171112111	DOCOMENTS		2013/0288252	A1	10/2013	Sparks et al.
2010/0291572 A1		Stoughton et al.		2013/0303461 2013/0310260			Iafrate et al. Kim et al.
2010/0291635 A1 2010/0323352 A1	11/2010			2013/0310200			Lo et al.
2010/0323332 A1 2011/0015096 A1	1/2010	Lo et al. Chiu		2013/0325360			Deciu et al.
2011/0033862 A1		Rabinowitz et al.		2014/0032128			Rabinowitz et al.
2011/0039724 A1		Lo et al.		2014/0038830 2014/0051585		2/2014	Srinivasan et al. Prosen et al.
2011/0045462 A1 2011/0064824 A1		Fu et al. Lascoste et al.		2014/0065621			Mhatre et al.
2011/0004824 A1 2011/0071031 A1		Khripin et al.		2014/0066317	A1	3/2014	Talasaz
2011/0086769 A1	4/2011	Olipĥant et al.		2014/0087385			Rabinowitz et al.
2011/0092763 A1		Rabinowitz et al.		2014/0094373 2014/0100126			Zimmermann et al. Rabinowitz
2011/0105353 A1 2011/0111410 A1		Lo et al. Ryan et al.		2014/0100134		4/2014	Rabinowitz et al.
2011/0130558 A1		Ritt et al.		2014/0106975			Stoughton et al.
2011/0151442 A1		Fan et al.		2014/0113795 2014/0141981			Emerson et al. Zimmermann et al.
2011/0159499 A1 2011/0160078 A1		Hindson et al. Fodor et al.		2014/0154682			Rabinowitz et al.
2011/0100078 A1 2011/0178719 A1		Rabinowitz et al.		2014/0155274	A1	6/2014	Xie et al.
2011/0201507 A1	8/2011	Rava et al.		2014/0162269			Rabinowitz
2011/0212446 A1		Wang et al.		2014/0186827 2014/0193816			Pieprzyk et al. Rabinowitz et al.
2011/0212846 A1 2011/0224087 A1	9/2011 9/2011	Ouake et al.		2014/0206552			Rabinowitz et al.
2011/0246083 A1		Fan et al.		2014/0227691			May et al.
2011/0251149 A1		Perrine et al.		2014/0227705			Vogelstein et al. Van Den Boom et al.
2011/0288780 A1 2011/0294699 A1		Rabinowitz et al.		2014/0242588 2014/0256558			Varley et al.
2011/0294699 A1 2011/0300608 A1		Lee et al. Ryan et al.		2014/0256569		9/2014	Rabinowitz et al.
2011/0301854 A1		Curry et al.		2014/0272956			Huang et al.
2011/0312503 A1		Chuu et al.		2014/0274740 2014/0287934		9/2014	Srinivasan et al. Szelinger et al.
2011/0318734 A1 2012/0003635 A1		Lo et al. Lo et al.		2014/0296081			Diehn et al.
2012/0003635 A1 2012/0003637 A1		Lo et al.		2014/0329245	A1	11/2014	Spier et al.
2012/0010085 A1		Rava et al.		2014/0336060			Rabinowitz
2012/0028814 A1	2/2012			2015/0051087 2015/0064695			Rabinowitz et al. Katz et al.
2012/0034603 A1 2012/0034685 A1		Oliphant et al. Sparks et al.		2015/0086477			Mitchell et al.
2012/0108460 A1	5/2012			2015/0087535		3/2015	Patel
2012/0122701 A1		Ryan et al.		2015/0099673 2015/0147815		4/2015	Fodor Babiarz et al.
2012/0165203 A1 2012/0185176 A1	6/2012	Quake et al. Rabinowitz et al.		2015/0147815		6/2015	
2012/0183176 A1 2012/0190020 A1		Oliphant et al.		2015/0197786			Osborne et al.
2012/0190021 A1	7/2012	Oliphant et al.		2015/0211050			Infrate et al.
2012/0190557 A1		Oliphant et al.		2015/0218631 2015/0232938			Chuu et al. Mhatre
2012/0191358 A1 2012/0196754 A1		Oliphant et al. Quake et al.		2015/0265995			Head et al.
2012/0208706 A1		Downing et al.		2015/0299812		10/2015	
2012/0214678 A1		Rava et al.		2015/0315657 2015/0322507			Rhodes et al. Zimmermann et al.
2012/0251411 A1 2012/0264121 A1	10/2012	Jeon Rava et al.		2015/0322891			Tan et al.
2012/0264618 A1	10/2012			2016/0032396		2/2016	Diehn et al.
2012/0270212 A1	10/2012	Rabinowitz et al.		2016/0046986			Eltoukhy et al.
2012/0270739 A1		Rava et al.		2016/0115541 2016/0145682			Schutz et al. Woodward et al.
2012/0295810 A1 2012/0295819 A1		Quake et al. Leamon et al.		2016/0186239		6/2016	Sinha
2013/0017549 A1	1/2013			2016/0186253		6/2016	Talasaz et al.
2013/0022973 A1		Hansen et al.		2016/0201124 2016/0239602			Donahue et al. Shendure et al.
2013/0024127 A1 2013/0034546 A1	1/2013 2/2013			2016/0239002			Babiarz et al.
2013/0034340 A1 2013/0040375 A1	2/2013			2016/0257993	A1	9/2016	Fu et al.
2013/0060483 A1	3/2013	Struble et al.		2016/0265042			Schroeder et al.
2013/0069869 A1	3/2013			2016/0289740 2016/0289753			Fu et al. Osborne et al.
2013/0085681 A1 2013/0090250 A1	4/2013	Deciu et al. Sparks et al.		2016/0312276			Fu et al.
2013/0030230 A1 2013/0116130 A1		Fu et al.		2016/0319345		11/2016	Gnerre et al.
2013/0123120 A1	5/2013	Zimmermann et al	l.	2016/0369333			Babiarz et al.
2013/0130923 A1		Ehrich et al.		2017/0011166 2017/0107576			Rabinowitz et al. Babiarz et al.
2013/0157870 A1 2013/0172211 A1	6/2013 7/2013	Pushkarev et al. Oliphant et al.		2017/0107376			Mitchell et al.
2013/0172211 A1 2013/0178373 A1	7/2013			2017/0121716			Rodi et al.
2013/0190653 A1	7/2013	Alvarez Ramos		2017/0145475	A1	5/2017	Hunsley et al.
2013/0196862 A1	8/2013			2017/0152561			Hamamah et al.
2013/0210644 A1 2013/0225422 A1	8/2013 8/2013			2017/0218458 2017/0275689			Fan et al. Maguire et al.
2013/0223422 A1 2013/0237431 A1		Lo et al.		2017/02/3089			Khoja et al.
2013/0257431 A1 2013/0252824 A1		Rabinowitz		2017/0203700			Green et al.

(56)	References Cited	2020/0126634		2020	
HC	DATENT DOCUMENTS	2020/0140950 2020/014911			Babiarz et al. Babiarz et al.
0.8	PATENT DOCUMENTS	2020/014911			Babiarz et al.
2017/0342477 A1	11/2017 Jensen et al.	2020/017297			Rabinowitz et al.
2017/0342477 A1 2017/0362649 A1	12/2017 Jensen et al. 12/2017 Lieberman-Aide	2020/010100			Rabinowitz et al.
2018/0023128 A1	1/2018 Yanai et al.	2020/0190570			Ryan et al.
2018/0025109 A1	2/2018 Rabinowitz et al	. 2020/0190573			Rabinowitz et al.
2018/0127744 A1	5/2018 Hu et al.	2020/0190593			Rabinowitz et al.
2018/0148777 A1	5/2018 Kirkizlar et al.	2020/0208196 2020/020822			Zimmermann et al. Babiarz et al.
2018/0155775 A1	6/2018 Zimmermann et				Rabinowitz et al.
2018/0155776 A1 2018/0155779 A1	6/2018 Zimmermann et 6/2018 Zimmermann et	2020/022202			Rabinowitz et al.
2018/0155785 A1	6/2018 Rabinowitz et al	ai. 2020/022202			Babiarz et al.
2018/0155786 A1	6/2018 Rabinowitz et al	. 2020/0248264			Rabinowitz et al.
2018/0155792 A1	6/2018 Rabinowitz et al	. 2020/0248266		2020	
2018/0171409 A1	6/2018 Rabinowitz et al	. 2020/0316498 2020/031819			Mitchell Babiarz et al.
2018/0171420 A1	6/2018 Babiarz et al.	2020/02/47/45			Babiarz et al.
2018/0173845 A1 2018/0173846 A1	6/2018 Sigurjonsson et 6/2018 Sigurjonsson et	2020/025002			Rabinowitz et al.
2018/01/3840 A1 2018/0187241 A1	7/2018 Selvaraj et al.	2020/036241			Rabinowitz et al.
2018/0201995 A1	7/2018 Rabinowitz et al	2020/0407788			Ryan et al.
2018/0237841 A1	8/2018 Stray et al.	2020/0407798			Babiarz et al.
2018/0251553 A1	9/2018 McGranahan et	al. 2021/0009990		2021 2021	
2018/0265917 A1	9/2018 Barany et al.	2021/0025005 2021/0054459			Rabinowitz et al.
2018/0288982 A1	10/2018 Sinha 10/2018 Ryan et al.	2021/0071240			Zimmermann et al.
2018/0298439 A1 2018/0300448 A1	10/2018 Ryan et al. 10/2018 Rabinowitz et al	2021/012006			Mitchell et al.
2018/0320171 A1	11/2018 Withey	2021/0139983			Mitchell et al.
2019/0010543 A1	1/2019 Babiarz et al.	2021/0139988			Mitchell et al.
2019/0106737 A1	4/2019 Underhill	2021/0155988			Rabinowitz et al.
2019/0106751 A1	4/2019 Zimmermann et	al. 2021/0189498 2021/0198733		2021	Babiarz et al. Moshkevich et al.
2019/0112661 A1	4/2019 Khan et al.	2021/010074			Rabinowitz et al.
2019/0185913 A1 2019/0185936 A1	6/2019 Zimmermann et 6/2019 Babiarz et al.	ai. 2021/0198743			Rabinowitz et al.
2019/0194743 A1	6/2019 Ryan et al.	2021/0222230			Zimmermann et al.
2019/0194758 A1	6/2019 Babiarz et al.	2021/0222240			Moshkevich et al.
2019/0194759 A1	6/2019 Babiarz et al.	2021/0257048			Zimmermann et al.
2019/0203290 A1	7/2019 Rabinowitz et al	. 2021/0269879 2021/0324463			Mitchell et al. Rabinowitz et al.
2019/0203294 A1	7/2019 Babiarz et al.	2021/022772		2021	
2019/0211391 A1 2019/0211392 A1	7/2019 Rabinowitz et al 7/2019 Rabinowitz et al			2021	
2019/0211392 A1	7/2019 Rabinowitz et al	2021/0355536		2021	
2019/0211399 A1	7/2019 Rabinowitz et al	. 2022/0025455			Zimmermann et al.
2019/0211402 A1	7/2019 Babiarz et al.	2022/0025456			Rabinowitz et al.
2019/0211406 A1	7/2019 Babiarz et al.	2022/0033908 2022/0033909			Rabinowitz et al. Babiarz et al.
2019/0249241 A1 2019/0256894 A1	8/2019 Rabinowitz et al 8/2019 Zimmermann et	. 2022/004210			Rabinowitz et al.
2019/0256906 A1	8/2019 Ziffinermann et 8/2019 Rabinowitz et al	2022/005650			Zimmermann
2019/0256907 A1	8/2019 Ryan et al.	2022/0056534			Rivers
2019/0256908 A1	8/2019 Rabinowitz et al	. 2022/0073978			Rabinowitz et al.
2019/0256909 A1	8/2019 Rabinowitz et al			2022	
2019/0256912 A1	8/2019 Rabinowitz et al	. 2022/009866° 2022/013949			Rabinowitz et al. Rabinowitz et al.
2019/0256916 A1 2019/0256917 A1	8/2019 Babiarz et al. 8/2019 Babiarz et al.	2022/0154249			Zimmermann et al.
2019/0256919 A1	8/2019 Babiarz et al.	2022/0154290			Babiarz et al.
2019/0256924 A1	8/2019 Vogelstein et al.	2022/0195520		2022	Rabinowitz et al.
2019/0256931 A1	8/2019 Babiarz et al.	2022/021356			Babiarz et al.
2019/0264277 A1	8/2019 Rabinowitz et al				Hafez et al
2019/0264280 A1	8/2019 Rabinowitz et al 8/2019 Rabinowitz et al		A1 9/.	2022	Babiarz et al.
2019/0264288 A1 2019/0271043 A1	9/2019 Rabinowitz et al.		ODEICNI E	MTD.	NIT DOCUMENTO
2019/0276888 A1	9/2019 Rabinowitz et al	. FO	JKEIGN F	AIE.	NT DOCUMENTS
2019/0284623 A1	9/2019 Rabinowitz et al		101675169	Δ	3/2010
2019/0300950 A1	10/2019 Rabinowitz et al	· CN	104736722		6/2015
2019/0309358 A1	10/2019 Rabinowitz et al	· CN	105229175		1/2016
2019/0309359 A1 2019/0309365 A1	10/2019 Zimmermann et 10/2019 Babiarz et al.	EI	0270017		6/1988
2019/0309303 A1 2019/0316177 A1	10/2019 Zimmermann et	al. EP	1325963		7/2003
2019/0316184 A1	10/2019 Zimmermann et		1524321 1325963		4/2005 9/2006
2019/0316200 A1	10/2019 Rabinowitz et al	· EP	1524321		7/2009
2019/0323076 A1	10/2019 Rabinowitz et al	· EP	2163622		3/2010
2019/0360036 A1	11/2019 Rabinowitz et al	· EP	2128169	A1	12/2010
2019/0367972 A1	12/2019 Mitchell et al.	EP	2653562		10/2013
2020/0024653 A1 2020/0032323 A1	1/2020 Bethke 1/2020 Talasaz et al.	EP	2902500		8/2015
2020/0032323 A1 2020/0032340 A1	1/2020 Talasaz et al. 1/2020 Mitchell et al.	EP EP	3026124 2315849		6/2016 11/2017
2020/0109449 A1	4/2020 Stamm et al.	EP	3285193		2/2018
2020/0123612 A1	4/2020 Babiarz et al.	EP	2877594		12/2019

## US 11,530,454 B2

Page 6

		_		
(56)	References Cited	WO WO	2007/117256 A1 2007117039 A1	10/2007 10/2007
	FOREIGN PATENT DOCUMENTS	WO	2007132167 A2	11/2007
ED	2105505 D1 (/2020	WO WO	2007/147073 A2 2007/147076 A2	12/2007 12/2007
EP EP	3187597 B1 6/2020 3134541 B1 8/2020	WO	2007140417 A2	12/2007
EP	3760730 A1 1/2021	WO WO	2007147074 A2 2007147079 A2	12/2007 12/2007
EP EP	3760731 A1 1/2021 3760732 A1 1/2021	WO	2007147079 A2 2008024473 A2	2/2007
EP	3824470 5/2021	WO	2008048931 A1	4/2008
EP GB	3443119 B1 2/2022 2488358 8/2012	WO WO	2008/061213 A2 2008051928 A2	5/2008 5/2008
JP	2965699 8/1999	WO	2008056937 A1	5/2008
JP JP	2002-530121 A 9/2002 2002-300894 A 10/2002	WO WO	2008059578 A1 2008079374 A2	5/2008 7/2008
JР	2002-300894 A 10/2002 2003/521252 A 7/2003	WO	2008081451 A2	7/2008
JP	2004502466 A 1/2004	WO WO	2008084405 A2 2008115427 A2	7/2008 9/2008
JP JP	2004533243 A 11/2004 2005514956 A 5/2005	WO	2008115497 A2	9/2008
JP	2005160470 A 6/2005	WO WO	2008118988 A1 2008135837 A2	10/2008 11/2008
JP JP	2006-254912 A 9/2006 2008-263974 A 11/2008	WO	2008157857 A2 2008157264 A2	12/2008
JP	2008/271980 A 11/2008	WO WO	2009009769 A2 2009013492 A1	1/2009 1/2009
JP JP	2010-509922 A 4/2010 2011/508662 A 3/2011	WO	2009013492 A1 2009013496 A1	1/2009
JP	2011/516069 A 5/2011	WO	2009019215 A1 2009019455 A2	2/2009
JP RU	2015-535681 12/2015 2290078 C1 12/2006	WO WO	2009/032779 A2	2/2009 3/2009
WO	95/01796 1/1995	WO	2009/036525 A2	3/2009
WO WO	1996036736 A2 11/1996 98/39474 9/1998	WO WO	2009030100 A1 2009032781 A2	3/2009 3/2009
wo	98/44151 10/1998	WO	2009033178 A1	3/2009
WO WO	00/18957 4/2000 2001007640 A2 2/2001	WO WO	2009049889 A1 2009017784 A2	4/2009 5/2009
WO	2001007640 A2 2/2001 0134844 A1 5/2001	WO	2009064897 A2	5/2009
WO	01/57269 A2 8/2001	WO WO	2009091934 A1 2009092035 A2	7/2009 7/2009
WO WO	2001/079851 A1 10/2001 200190419 A2 11/2001	WO	2009/105531 A1	8/2009
WO	2002004672 A2 1/2002	WO WO	2009099602 A1 2009100029 A1	8/2009 8/2009
WO WO	02/44411 A1 6/2002 2002055985 A2 7/2002	wo	2009105029 A1 2009105531 A1	8/2009
WO	02/070751 A1 9/2002	WO WO	2009117122 A2 2009120808 A2	9/2009 10/2009
WO WO	2002076377 10/2002 02/090505 A2 11/2002	wo	2009120808 A2 2009145828 A2	12/2009
WO	03/000919 A2 1/2003	WO WO	2009146335 A1 2010014920 A1	12/2009 2/2010
WO WO	03/018757 A3 3/2003 03/020974 A3 3/2003	WO	2010014920 A1 2010017214 A1	2/2010
WO	2003031646 A1 4/2003	WO	2010/033639 A2	3/2010
WO WO	2003/050532 A1 6/2003 2003062441 A1 7/2003	WO WO	2010/033652 A1 2010033578 A2	3/2010 3/2010
wo	2003/102595 A1 12/2003	WO	2010042831 A2	4/2010
WO WO	2003/106623 A2 12/2003 2004/051218 A2 6/2004	WO WO	2010045617 A2 2010075459	4/2010 7/2010
wo	WO-2004051218 A2 * 6/2004 C12Q 1/6806	WO	2010/088288 A2	8/2010
WO WO	2004069849 A2 8/2004 2004070005 A2 8/2004	WO WO	2010115016 A2 2010115154 A1	10/2010 10/2010
WO	2004070005 A2	WO	2010118016 A2	10/2010
WO	2004087863 A2 10/2004	WO WO	2010/127186 A1 2011/023078 A1	11/2010 3/2011
WO WO	2005003375 A2 1/2005 2005021793 A1 3/2005	WO	2011/032078 A1	3/2011
WO	2005023091 A2 3/2005	WO WO	2011041485 A1 2011/051283 A1	4/2011 5/2011
WO WO	2005030999 A1 4/2005 2005035725 A2 4/2005	WO	2011/057061 A1	5/2011
WO	2005/039389 A3 5/2005	WO WO	2011057094 2011/090556 A1	5/2011 7/2011
WO WO	2005100401 A2 10/2005 2005123779 A2 12/2005	WO	2011/090330 A1 2011087760	7/2011
WO	2007145612 A1 6/2006	WO	2011102998 A2	8/2011
WO WO	2006/10855 A2 10/2006 2006/128192 A2 11/2006	WO WO	2011/118603 2011109440 A1	9/2011 9/2011
WO	2007/011903 A3 1/2007	WO	2011140433 A2	11/2011
WO WO	2007/052006 A1 5/2007 2007057647 A1 5/2007	WO WO	2011146632 A1 2012/019200 A2	11/2011 2/2012
WO	2007062164 A3 5/2007	WO	2012/019200 A2 2012/028746 A1	3/2012
WO	2007/073171 A2 6/2007	WO	2012042374 A2	4/2012
WO WO	2007070280 A2 6/2007 2007070482 A2 6/2007	WO WO	2012/058488 A1 201283250	5/2012 6/2012
WO	2007075836 A2 7/2007	WO	2012088456 A2	6/2012
WO	2007/092473 A2 8/2007 2007086935 A2 8/2007	WO WO	20120071621 2012108920 A1	6/2012 8/2012
WO	2007086935 A2 8/2007	WO	2012108920 Al	0/2012

(56)	References Cited	OTHER PUBLICATIONS
	FOREIGN PATENT DOCUMENTS	Marusyk et al. Tumor heterogeneity: Causes and consequences
WO	2012/142531 A2 10/2012	Biochimica et Biophysics Acts vol. 1805, pp. 105-117 (Year: 2010).*
WO WO	2007/149791 A2 12/2012 2013030577 3/2013	Diehl et al. Detection and quantification of mutations in the plasma
wo	2013/045432 A1 4/2013	of patients with colorectal tumors Proceedings of the National
WO	2013/049892 A1 4/2013	Academy of Sciences USA vol. 102 pp. 16368-16373 (Year: 2005).*
WO WO	2013052557 A2 4/2013 2013/078470 A2 5/2013	Ku et al. Exome versus transcriptome sequencing in identifying
wo	2013/086464 A1 6/2013	coding region variants Expert Review of Molecular Diagnostics vol. 12, pp. 241-251 (Year: 2012).*
WO	2013/123220 A1 8/2013	European Application No. 014198110, European Search Report
WO WO	2013/138510 A1 9/2013 2013/138510 A9 9/2013	dated Apr. 28, 2015, 3 pages.
WO	20130130848 9/2013	PRNewswire (Research Suggests Daily Consumption of Orange
WO WO	2013/159035 A2 10/2013 2013/169339 A1 11/2013	Juice Can Reduce Blood Pressure and May Provide Beneficial
wo	2013/177220 A1 11/2013	Effects to Blood Vessel Function: New Study Identified Health Benefits in Orange Juice, Dec. 8, 2010), 3 pages.
WO	2013/181651 A1 12/2013	"Abstracts for CNAPS III Circulating Nucleic Acids in Plasma and
WO WO	2014/004726 A1 1/2014 2014/014497 A1 1/2014	Serum and Serum Proteomics", Clinical Chemistry, vol. 49, No. 11,
wo	20140018080 1/2014	2003, 33 pages.
WO	2014/035986 A1 3/2014	"Abstracts for CNAPS IV Circulating Nucleic Acids in Plasma/
WO WO	2014/122288 A1 8/2014 2014/145078 A1 9/2014	Serum", Fourth International Conference on Circulating Nucleic Acids in Plasma/Serum (CNAPS-IV), 2005, 40 pages.
WO	2014/145232 A2 9/2014	"How Many Carbs in a Potato?, [Online]", New Health Guide, Nov.
WO WO	2014/149134 A2 9/2014 2014/150300 A2 9/2014	1, 2014, 3 pages.
WO	2014/150300 A2 9/2014 2014/151117 A1 9/2014	"Random variable", The Penguin Dictionary of Mathematics, 2008,
WO	2014/194113 A2 12/2014	1 page.
WO WO	2015134552 A1 3/2015 2015/048535 A1 4/2015	"Fixed Medium", "Academic Press", http://www.xreferplus.com/entry.do?pp=1&id=310, 1996, 1 pg.
WO	2015/070086 A1 5/2015	Abaan, O. D et al., "The Exomes of the NCI-60 Panel: A Genomic
WO	2015/100427 A1 7/2015	Resource for Cancer Biology and Systems Pharmacology", Cancer
WO WO	2015/148494 A1 10/2015 2015/164432 A1 10/2015	Res., vol. 73, No. 14, Jul. 15, 2013, 4372-4382.
WO	2016/009059 A1 1/2016	Abbosh, C. et al., "Phylogenetic ctDNA analysis depicts early-stage
WO WO	2016009224 A1 1/2016 2016/063122 A1 4/2016	Tung cancer evolution", Nature, vol. 545, May 25, 2017, 446-451. Abd-Elsalam, Kamel A., "Bioinformatic Tools And Guideline for
wo	2016/065295 A1 4/2016	PCR Primer Design", African Journal of Biotechnology, vol. 2,
WO	2016/077313 A1 5/2016	2003, pp. 91-95.
WO WO	2016/123698 A1 8/2016 2016/138080 A1 9/2016	Abidi, S. et al., "Leveraging XML-based electronic medical records
WO	2016/176662 A1 11/2016	to extract experiential clinical knowledge: An automated approach to generate cases for medical case-based reasoning systems", Inter-
WO WO	2016/183106 A1 11/2016 2016/193490 A1 12/2016	national Journal of Medical Informatics, 68(1-3), 2002, 187-203.
wo	2017/058784 A1 4/2017	Adalsteinsson, V. A. et al., "Scalable whole-exome sequencing of
WO	2017/181146 A1 10/2017	cell-free DNA reveals high concordance with metastatic tumors",
WO WO	2017/181202 A2 10/2017 2017/190106 A1 11/2017	Nature Communications, vol. 18, No. 1324, 2017, 13 pages. Adinolfi, M. et al., "Rapid Detection of Aneuploidies by Microsatel-
WO	2017205540 A1 11/2017	lite and the Quantitative Fluorescent Polymerase Chain Reaction",
WO WO	2018/009723 A1 1/2018 2018/083467 A1 5/2018	Prenatal Diagnosis, vol. 17, No. 13, 1997, 1299-1311.
wo	2018/085603 A1 5/2018	Agarwal, Ashwin. et al., "Commercial Landscape of Noninvasive Prenatal Testing in the United States", Prenatal Diagnosis, 33, 2013,
WO	2018/106798 A1 6/2018	521-531.
WO WO	2018/136562 A2 7/2018 2018/156418 A1 8/2018	Agbor-Enoh, S. et al., "Donor-derived cell-free DNA predicts
WO	2018/237081 A1 12/2018	allograft failure and mortality after lung transplantation", EBioMedicine,
WO WO	2019/046817 A1 3/2019 2019/118926 A1 6/2019	vol. 40, 2019, 541-553.  Ahmadian, A. et al., "Analysis of the p53 Tumor Suppressor Gene
WO	2019/118920 A1 0/2019 2019/140298 A1 7/2019	by Pyrosequencing", BioTechniques, vol. 28, Jan. 2000, 140-147.
WO	2019/161244 A1 8/2019	Alaeddini, R. et al., "Forensic implications of genetic analyses from
WO WO	2019/200228 A1 10/2019 2019/241349 A1 12/2019	degraded DNA—A review", Forensic Science International: Genetics, vol. 4, 2010, 148-157.
wo	2020/010255 A1 1/2020	Alberts, B. et al., "Chapter 20: Germ Cells and Fertilization",
WO	2020/018522 A1 1/2020 2020/041440 A1 2/2020	Molecular Biology of the Cell, Fourth Edition, 2002, 1127-1156.
WO WO	2020/041449 A1 2/2020 2020/076957 A1 4/2020	Alberts, B. et al., "Chapter 4: DNA and Chromosomes", Molecular
WO	2020/106987 A1 5/2020	Biology of the Cell, Fourth Edition, 2002, 191-234. Alizadeh, Mehdi et al., "Quantitative Assessment of Hematopoietic
WO WO	2020104670 A1 5/2020 2020/131699 A2 6/2020	Chimerism after Bone Marrow Transplantation by Real-time Quan-
WO	2020/131099 A2 0/2020 2020/214547 A1 10/2020	titative Polymerase Chain Reaction", Blood, vol. 99, No. 12, Jun.
WO	2020/247263 A1 12/2020	15, 2002, 4618-4625.
WO WO	2021/055968 A1 3/2021 2007100911 A2 9/2021	Alkan, Can et al., "Personalized Copy Number and Segmental Duplication Maps Using Next-Generation Sequencing", Nature
wo	2007100911 A2 9/2021 2021/243045 A1 12/2021	Genetics, 41, 10, 2009, 1061-1068.
WO	2022/015676 A1 1/2022	Allaire, F.R., "Mate selection by selection index theory", Theo-
WO	2022197864 9/2022	retical Applied Genetics, 57(6), 1980, 267-272.

Case: 24-1324 Document: 42-1 Page: 322 Filed: 03/18/2024

### US 11,530,454 B2

Page 8

#### (56)References Cited

### OTHER PUBLICATIONS

Allan, J. et al., "Micrococcal Nuclease Does Not Substantially Bias Nucleosome Mapping", Journal of Molecular Biology, vol. 417, Jan. 30, 2012, 152-164.

Allawi, Hatim T. et al., "Thermodynamics of internal C•T Mismatches in DNA", Nucleic Acids Research, 26 (11), 1998, 2694-

Ambardar, S. et al., "High Throughput Sequencing: An Overview of Sequencing Chemistry", Indian J. Microbiol., vol. 56, No. 4, 2016,

Amicucci, P. et al., "Prenatal Diagnosis of Myotonic Dystrophy Using Fetal DNA Obtained from Maternal Plasma", Clinical Chem-

istry, vol. 46, No. 2, 2000, 301-302. Andras, S. C. et al., "Strategies for Signal Amplification in Nucleic Acid Detection", Molecular Biotechnology, vol. 19, 2001, 29-44. Anker, P. et al., "Circulating DNA in Plasma or Serum", Medicina, vol. 60, 2000, 699-702.

Anker, P. et al., "Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients", Cancer and Metastasis Reviews, vol. 18, 1999, 65-73.

Anker, P. et al., "The Second International Symposium on Circulating Nucleic Acids in Plasma and Serum (CNAPS-2) held in conjunction with the 6th Annual Scientific Symposium of the Hong Kong Cancer Institute", Clinical Chemistry, vol. 47, No. 2, 2001,

Ansorge, Wilhelm J., "Next-generation DNA Sequencing Techniques", New Biotechnology, vol. 25, No. 4, Feb. 2, 2009, 195-203. Antonarakis, S. E. et al., "Chromosome 21 and Down Syndrome: From Genomics to Pathophysiology", Nature Reviews Genetics, vol. 5, Oct. 2004, 725-738.

Aoki, Yasuhiro, "Statistical and Probabilistic Bases of Forensic DNA Testing", The Journal of the Iwate Medical Association, 2002, vol. 54, p. 81-94.

Arandjelovic, M. et al., "Two-Step Multiplex Polymerase Chain Reaction improves the Speed and Accuracy of Genotyping Using DNA from Noninvasive and Museum Samples", Molecular Ecology Resources, vol. 9, 2009, pp. 28-36.

Ashoor, G. et al., "Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics", Ultrasound in Obstetrics and Gynecology, vol. 41, 2013,

Ashoor, Ghalia et al., "Chromosome-Selective Sequencing of Maternal Plasma Cell-Free DNA for First-Trimester Detection of Trisomy 21 and Trisomy 18", American Journal of Obstetrics & Gynecology, 206, 2012, 322.e1-322.e5

Ashoor, Ghalia et al., "Fetal Fraction in Maternal Plasma Cell-Free DNA at 11-13 Weeks' Gestation: Effect of Maternal and Fetal Factors", Fetal Diagnosis Therapy, 2012, 1-7

Auld, D. S., "Use of Chelating Agents to Inhibit Enzymes", Methods in Enzymology, vol. 158, 1988, 110-114. Avent, Neil D. et al., "Cell-free Fetal DNA in the Maternal Serum

and Plasma: Current and Evolving Applications", Current Opinion in Obstretrics and Gynecology, vol. 21, No. 2, Apr. 1, 2009,

Avgidou, K. et al., "Prospective first-trimester screening for trisomy 21 in 30,564 pregnancies", American Journal of Obstetrics and Gynecology, vol. 192, 2005, 1761-1767. Ayala, et al., "Long-Term Follow-Up of Donor Chimerism Toler-

ance After Human Liver Transplantation", Liver Transplantation, vol. 15, No. 6, May 28, 2009, 581-591.

Bada, Michael A. et al., "Computational Modeling of Structural Experimental Data", Methods in Enzymology, 317, 2000, 470-491. Bai, H. et al., "Detection and Clinical Significance of Intratumoral EGFR Mutational Heterogeneity in Chinese Patients with Advanced Non-Small Cell Lung Cancer", PLOS One, vol. 8, No. 2, Feb. 2013, 7 pages

Balavoine, Guillaume, "Identification of Members of Several Homeobox Genes in a Planarian Using a Ligation-Mediated Polymerase Chain Reaction Technique", Nucleic Acids Research, vol. 24, 1996, pp. 1547-1553.

Balduini, et al., "Utility of Biochemical Markers in The Follow-up Heart Transplant Recipients", Transplantation Proceedings, vol. 35, No. 8, Dec. 1, 2003, 3075-3078.

Bale, J. R. et al., "Reducing Birth Defects: Meeting the Challenge in the Developing World", Institute of Medicine of the National Academies, 2003, 270 pgs.

Ballif, B. C. et al., "Detection of Low-Level Mosaicism by Array CGH in Routine Diagnostic Specimens", American Journal of Medical Genetics Part A, vol. 140A, 2006, 2757-2767.

Banfi, G. et al., "The role of ethylenediamine tetraacetic acid (EDTA) as in vitro anticoagulant for diagnostic purposes", Clin. Chem., vol. 45, No. 5, 2007, 565-576.

Barbazuk, et al., "SNP Discovery via 454 Transcriptome Sequencing", The Plant Journal, vol. 51, Jul. 27, 2007, 910-918.

Barra, G. B. et al., "EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples", Clinical Biochemistry, vol. 48, 2015, 976-981.

Barski, A. et al., "High-Resolution Profiling of Histone Methylations in the Human Genome", Cell, vol. 129, May 18, 2007, 823-837. Bartlett, John M. et al., "PCR Protocols", PCR Protocols, vol. 226, 2003, 519 pages.

Bashashati, A. et al., "Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling", Journal of Pathology, vol. 231, 2013, 21-34.

Bau, Stephan et al., "Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays", Anal Bioanal Chern, vol. 393, 2009, 171-175

Bauer, M. et al., "A prospective analysis of cell-free fetal DNA concentration in maternal plasma as an indicator for adverse pregnancy outcome", Prenatal Diagnosis, vol. 26, 2006, 831-836.

Baxter, L. L. et al., "Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse", Human Molecular Genetics, vol. 9, No. 2, Jan. 2000, 195-202. Baxter-Lowe, et al., "Tracking Microchimeric DNA In Plasma to

Diagnose and Manage Organ Transplant Rejection", Clinical Chemistry, vol. 52, No. 4, Apr. 1, 2006, 559-561. Beaumont, Mark A et al., "The Bayesian Revolution in Genetics",

Nature Reviews Genetics, 5, 2004, 251-261.

Beck, et al., "Next Generation Sequencing of Serum Circulating Nucleic Acids from Patients with Invasive Ductal Breast Cancer Reveals Differences to Healthy and Nonmalignant Controls", Molecular Cancer Research, vol. 8, No. 3, Mar. 1, 2010, 335-342

Beck, J. et al., "Digital Droplet PCR for Rapid Quantification of Donor DNA in the Circulation of Transplant Recipients as a Potential Universal Biomarker of Graft Injury", Clinical Chemistry, vol. 59, No. 12, 2013, 1732-1741. Beck, J. et al., "Profile of the Circulating DNA in Apparently

Healthy Individuals", Clinical Chemistry, vol. 55, No. 4, 2009,

Beer, Alan E. et al., "The Biological Basis of Passage of Fetal Cellular Material into the Maternal Circulation", Annals New York Academy of Sciences, 731, 1994, 21-35.
Beerenwinkel, et al., "Methods for Optimizing Antiviral Combina-

tion Therapies", Bioinformatics, 19(1), 2003, i16-i25.
Beerenwinkel, N. et al., "Geno2pheno: estimating phenotypic drug

resistance from HIV-1 genotypes", Nucleic Acids Research, 31(13), 2003, 3850-3855.

Belostotsky, Dmitry A. et al., "Plant Systems Biology", Methods in Molecular Biology, vol. 553, Aug. 25, 2009, 3-408

Bender, et al., "A Multiplex SNP Typing Approach for the DNA Pyrosequencing Technology", International Congress Series, vol. 1288, Apr. 20, 2006, 73-75

Benjamini, Y. et al., "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing", Journal of the Royal Statistical Society, Series B (Methodological), vol. 5 7, No. 1, 1995, 289-300.

Benn, P. et al., "Non-Invasive Prenatal Testing for Aneuploidy: Current Status and Future Prospects", Ultrasound Obstet Gynecol,

42, 2013, 15-33.
Benn, P. et al., "Non-Invasive prenatal Diagnosis for Down Syndrome: the Paradigm Will Shift, but Slowly", Ultrasound Obstet. Gynecol., 39, 2012, 127-130.

### US 11,530,454 B2

Page 9

### (56) References Cited

### OTHER PUBLICATIONS

Bennett, S. T. et al., "Toward the \$1000 human genome", Pharmacogenomics, vol. 6, No. 4, 2005, 373-382.

Bentley, et al., "High-resolution, High-throughput HLA Genotyping by Next-generation Sequencing", Tissue Antigens, vol. 74, No. 5, Nov. 1, 2009, 393-403.

Bentley, David R. et al., "Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry", Nature, 456, 6, 2008, 53-50

Bergen, A. W. et al., "Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance", BMC Biotechnology, vol. 5, No. 24, Sep. 16, 2005, 11 pgs. Bermudez, M. et al., "Single-cell sequencing and mini-sequencing

Bermudez, M. et al., "Single-cell sequencing and mini-sequencing for preimplantation genetic diagnosis", Prenatal Diagnosis, 23, 2003, 669-677.

Beroud, C. et al., "Prenatal diagnosis of spinal muscular atrophy by genetic analysis of circulating fetal cells", The Lancet, vol. 361, Mar. 22, 2003, 1013-1014.

Bevinetto, Gina, Bevinetto (5 Foods All Pregnant Women Need, American Baby, available at http://www.parents.com/pregnancy/mybody/nutrition/5greatpregnancyfoods/, Apr. 15, 2008), 8 pgs. Bianchi, D W. et al., "Fetal gender and aneuploidy detection using fetal cells maternal blood: analysis of NIFTY I data", Prenat Diagn 2002; 22, 2002, 609-615.

Bianchi, D.W. et al., "Insights Into Fetal and Neonatal Development Through Analysis of Cell-Free RNA in Body Fluids", Early Human Development, vol. 86, No. 11, Nov. 2010, 747-752.

Bianchi, D. W., "Circulating Fetal DNA: Its Origin and Diagnostic Potential—A Review", Placenta, vol. 25, Supplemental A, May 2004, S93-S101.

Bianchi, D. W. et al., "Genome-Wide Fetal Aneuploidy Detection by Maternal Plasma DNA Sequencing", Obstetrics & Gynecology, vol. 119, No. 5, May 2012, 890-901.

Bianchi, D. W., "Review: Fetal Cells in the Maternal Circulation: Feasibility for Prenatal Diagnosis", British Journal of Haematology, vol. 105, 1999, 574-583.

Binladen, J. et al., "The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing", PLOS One, Issue 2, Feb. 2007, 9 pages.

Birch, Lyndsey et al., "Accurate and Robust Quantification of Circulating Fetal and Total DNA in Maternal Plasma from 5 to 41 Weeks of Gestation", Clinical Chemistry, 51(2), 2005, 312-320. Birkenkamp-Demtroder, et al., "Longitudinal assessment of multi-

plex patient-specific ctDNA biomarkers in bladder cancer for diagnosis, surveillance and recurrence", Annals of Oncology, Oxford University Press NLD, vol. 29, No. Supplement 8, 2018, viii26. Birkenkamp-Demtroder, K. et al., "Abstract 3653: Sequencing of

Birkenkamp-Demtroder, K. et al., "Abstract 3653: Sequencing of plasma cfDNA from patients with locally advanced bladder cancer for surveillance and therapeutic efficacy monitoring", Cancer Research, vol. 78, No. 13 Supplement, Jul. 2019, 1 page.

Bischoff, F. Z. et al., "Cell-free fetal DNA in maternal blood: kinetics, source and structure", Human Reproduction Update, vol. 11, No. 1, 2005, 59-67.

Bischoff, F. Z. et al., "Intact fetal cells in maternal plasma: are they really there?", Lancet, vol. 361, 2003, 139-140.

Bisignano, et al., "PGD and Aneuploidy Screening for 24 Chromosomes: Advantages and Disadvantages of Competing Platforms", Reproductive BioMedicine Online, 23, 2011, 677-685.

Blomquist, T M. et al., "Targeted RNA-Sequencing with Competitive Multiplex—PCR Amplicon Libraries", Pios One, vol. 8, Issue 11, Nov. 2013, 14 pages.

Blow, N., "The personal side of genomics", Nature, vol. 449, Oct. 4, 2007, 627-630.

Board, R.E. et al., "Detection of BRAF mutations in the tumour and serum of patients enrolled in the AZD6244 (ARRY-142886) advanced melanoma phase II study", British Journal of Cancer, vol. 101, 2009, 1724-1730.

Bodenreider, O., "The Unified Medical Language System (UMLS) Integrating Biomedical Terminology", Nucleic Acids Research, 32, (Database issue), 2004, D267-D270.

Bolotin, D. A. et al., "MiXCR: software for comprehensive adaptive immunity profiling", Nature, vol. 12, No. 5, May 2015, 380-381. Bordoni, et al., "Evaluation of Human Gene Variant Detection in Amplicon Pools by the GS-FLX Parallel Pyrosequencer", BMC Genomics, vol. 9, Oct. 8, 2008, 1-8.

Genomics, vol. 9, Oct. 8, 2008, 1-8. Boudsocq, F et al., "Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archael DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poln", Nucleic Acids Research, vol. 29, No. 22, 2001, 4607-4616.

29, No. 22, 2001, 4607-4616.

Bouma, B. N. et al., "Human Blood Coagulation Factor", The Journal of Biological Chemistry, vol. 252, No. 18, 1977, 6432-6437. Brastianos, P. K. et al., "Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets", Cancer Discovery, vol. 5, Sep. 26, 2015, 1164-1177.

Breithaupt, Holger, "The Future of Medicine", EMBO Reports, 21(61), 2001, 465-467.

Brinza, D. et al., "2SNP: scalable phasing based on 2-SNP haplotypes", Bioinformatics, vol. 22, No. 3, 2006, 371-373.

Brochet, X. et al., "IMGT/V-QUEST: the highly customized and

Brochet, X. et al., "IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis", Nucleic Acids Research, vol. 36, May 23, 2008, W503-W508.

Brockman, et al., "Quality Scores and SNP Detection in Sequencingby-synthesis Systems", Genome Research, vol. 18, No. 5, May 1, 2008, 763-770.

Broude, N E. et al., "High-Level Multiplex DNA Amplification", Antisense & Nucleic Acid Drug Development, vol. 11, 2001, 327-332.

Broude, N. E. et al., "High Multiplexity PCR Based on PCR Suppression", DNA Amplification Current Technologies and Applications, 2004, 61-76.

Broude, N. E. et al., "Multiplex Allele-specific Target Amplification based on PCR Suppression", PNAS, vol. 98, No. 1, Jan. 2, 2001, 206-211

Brownie, Jannine et al., "The Elimination of Primer-Dimer Accumulation in PCR", Nucleic Acids Research, 25(16), 1997, 3235-3241

Browning, S. R. et al., "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering", The American Journal of Human Genetics, vol. 81, Nov. 2007, 1084-1097.

Bryant, A. P., "Terminology of Sugars", Ind. Eng. Chem., vol. 26, No. 2, 1933, 231.

Bunnapradist, S. et al., "Using both the fraction and Quantity of Donor-Derived Cell-free DNA to Detect Kidney Allograft Rejection", JASN, vol. 32, 2021, 2439-2441.
Burkey, B. F. et al., "Hepatic apolipoprotein J is secreted as a

Burkey, B. F. et al., "Hepatic apolipoprotein J is secreted as a Tipoprotein", Journal of Lipid Research, vol. 33, 1992, 1517-1526. Burkova, E. E. et al., "Extremely Stable Soluble High Molecular Mass Multi-Protein Complex with DNase Activity in Human Placental Tissue", PLOS One, vol. 9, No. 11: e011234, Nov. 26, 2014, 26 pages.

Burnham, P. et al., "Myriad Applications of Circulating Cell-Free DNA in Precision Organ Transplant Monitoring", Annals of the American Thoracic Society, vol. 14, Supplement 3, Sep. 2017, S237-S241.

Burnham, P. et al., "Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma", Scientific Reports, vol. 6, No. 27859, Jun. 14, 2016, 9 pages.

Bustamante-Aragones, Ana et al., "New Strategy for the Prenatal Detection/Exclusion of Paternal Cystic Fibrosis Mutations in Maternal Plasma", Journal of Cystic Fibrosis, vol. 7, Issue 6, Nov. 1, 2008, 505-510.

Butler, et al., "Cardiovascular Magnetic Resonance in the Diagnosis of Acute Heart Transplant Rejection: A Review", Journal of Cardiovascular Magnetic Resonance, vol. 11, No. 1, Mar. 12, 2009, 1-11.

Butler, J. et al., "The Development of Reduced Size STR Amplicons as Tools for Analysis of Degraded DNA\*", Journal of Forensic Sciences, vol. 48, No. 5, 2003, 1054-1064.

### US 11,530,454 B2

Page 10

### (56) References Cited

### OTHER PUBLICATIONS

Butt, A. N. et al., "Overview of Circulating Nucleic Acids in Plasma/Serum: Update on Potential Prognostic and Diagnostic Value in Diseases Excluding Fetal Medicine and Oncology", Ann. N.Y. Acad. Sci., vol. 1137, 2008, 236-242.

Cairns, Paul et al., "Homozygous Deletions of 9p21 in Primary Human Bladder Tumors Detected by Comparative Multiplex Polymerase Chain Reaction", Cancer Research, 54, 1994, 1422-1424. Caliendo, Angela, "Multiplex PCR and Emerging Technologies for the Detection of Respiratory Pathogens", Clinical Infectious Diseases, 52(4), 2011,5326-S330.

Calin, G. A. et al., "A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia", N Engl J Med, vol. 353, 2005, 1793-1801.

Campbell, P. J. et al., "Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing", PNAS, vol. 105, No. 35, Sep. 2, 2008, 13081-13086.

2, 2008, 13081-13086.
Canick, J. A. et al., "The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies", Prenatal Diagnosis, vol. 33, 2013, 667-674.

Cansar, "Hs-578-T—Copy Number Variation—Cell Line Synopsis", ICR Cancer Research UK, Retrieved on Mar. 26, 2018 from https://cansar.icr.ac.uk/cansar/cell-lines/Hs-578-T/copy\_number\_variation/chromosome\_8/, Mar. 26, 2018, 50 pgs.

Cao, Y. et al., "Clinical Evaluation of Branched DNA Signal Amplification for Quantifying HIV Type 1 in Human Plasma", AIDS Research and Human Retroviruses, vol. 11, No. 3, 1995, 353-361.

Carnevale, Alessandra et al., "Attitudes of Mexican Geneticists Towards Prenatal Diagnosis and Selective Abortion", American Journal of Medical Genetics, 75, 1998, 426-431.

Carvalho, B. et al., "Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data", Biostatistics, vol. 8, No. 2, 2007, 485-499.

Casbon, J. A. et al., "A method for counting PCR template molecules with application to next-generation sequencing", Nucleic Acids Research, vol. 39, No. 12, Apr. 13, 2011, 1-8.

Castleberry, C. D. et al., "Quantification of Circulating Cell-Free DNA in Pediatric Heart Transplant Recipients", Journal of Heart and Lung Transplantation, vol. 30, No. 4, Apr. 1, 2011, S139.

Cawkwell, L. et al., "Rapid detection of allele loss in colorectal tumours using microsatellites and fluorescent DNA technology", Br. J. Cancer, vol. 67, 1993, 1262-1267.

Chakraborty, R. et al., "Paternity Exclusion by DNA Markers: Effects of Paternal Mutations", Journal of Forensic Sciences, vol. 41, No. 4, Jul. 1996, 671-677. Chan, Allen K. et al., "Cell-free Nucleic Acids In Plasma, Serum

Chan, Allen K. et al., "Cell-free Nucleic Acids In Plasma, Serum and Urine: A New Tool In Molecular Diagnosis", Annals of Clinical Biochemistry, vol. 40, Issue 2, Mar. 1, 2003, 122-130.

Chan, K.C. et al., "Size Distributions of Maternal and Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 50, No. 1, 2004, 88-92. Chang, H.W. et al., "Assessment of Plasma DNA Levels, Allelic Imbalance, and CA 125 as Diagnostic Tests for Cancer", Journal of the National Cancer Institute, vol. 94, No. 22, Nov. 20, 2002, 1697-1703.

Chavali, Sreenivas et al., "Oligonucleotide Properties Determination and Primer Designing: A Critical Examination of Predictions", Bioinformatics, vol. 21, 2005, pp. 3918-3925.

Chen, et al., "Non-invasive prenatal diagnosis using fetal DNA in maternal plasma: a preliminary study for identification of paternally-inherited alleles using single nucleotide polymorphisms", BMJ Open, 5(7), 2015, 1-8.

Chen, E. et al., "Noninvasive Prenatal Diagnosis of Fetal Trisomy 18 and Trisomy 13 by Maternal Plasma DNA Sequencing", PLoS ONE, 6(7), e21791, 2011, 7 pgs.

Chen, Bing-Yuan et al., "PCR Cloning Protocols", PCR Cloning Protocols, vol. 192, 2002, 434 pages.

Chen, C. P. et al., "Fetal DNA in maternal plasma: the prenatal detection of a paternally inherited fetal aneuploidy", Prenatal Diagnosis, vol. 20, 2000, 353-357.

Chen, X. Q. et al., "Microsatallite alterations in plasma DNA of small cell Tung cancer patients", Nature Medicine, vol. 2, No. 9, Sep. 1996, 1033-1035.

Chetty, Shilpa et al., "Uptake of Noninvasive Prenatal Testing (NIPT) in Women Following Positive Aneuploidy Screening", Prenatal Diagnosis, 33, 2013, 542-546.

Cheung, S. W. et al., "Rapid Publication: Microarray-Based CGH Detects Chromosomal Mosaicism Not Revealed by Conventional Cytogenetics", American Journal of Medical Genetics Part A, vol. 143A, 2007, 1679-1686.

Cheung, V. G. et al., "Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on Tess than one nanogram of genomic DNA", Proceedings of the National Academy of Sciences, USA, vol. 93, Dec. 1996, 14676-14679.

Chim, S. S. et al., "Detection and Characterization of Placental MicroRNAs in Maternal Plasma", Clinical Chemistry, vol. 54, No. 3, 2008, 482-490.

Chinnapapagari, S. K. et al., "Treatment of Maternal Blood Samples with Formaldehyde Does Not Alter the Proportion of Circulatory Fetal Nucleic Acids (DNA and mRNA) in Maternal Plasma", Clinical Chemistry, vol. 51, No. 3, 2005, 653-655.

Chitty, L. S. et al., "Noninvasive Prenatal Screening for Genetic Diseases Using Massively Parallel Sequencing of Maternal Plasma DNA", Cold Spring Harbor Perspectives in Medicine, vol. 5, No. 9, 2015, 20 pages.

Chiu, R. et al., "Non-Invasive Prenatal Assessment of Trisomy 21 by Multiplexed Maternal Plasma DNA Sequencing: Large Scale Validity Study", BMJ, 342, c7401, 2011, 9 pgs.

Chiu, R.W.K. et al., "Hypermethylation of RASSF1A in Human and Rhesus Placentas", The American Journal of Pathology, vol. 170, No. 3, Mar. 2007, 941-950.

Chiu, Rossa W. et al., "Effects of Blood-Processing Protocols on Fetal and Total DNA Quantification in Maternal Plasma", Clinical Chemistry, 47(9), 2001, 1607-1613.

Chiu, Rossa W.K. et al., "Maternal Plasma DNA Analysis with Massively Parallel Sequencing by Litigation for Noninvasive Prenatal Diagnosis of Trisomy 21", Clinical Chemistry, 56, 3, 2010, 459-463.

Chiu, Rossa W.K. et al., "Non-Invasive Prenatal Diagnosis by Single Molecule Counting Technologies", Trends in Genetics, 25 (7), 2009, 324-331.

Chiu, Rossa W.K. et al., "Noninvasive Prenatal Diagnosis of Fetal Chromosomal Aneuploidy by Massively Parallel Genomic Sequencing of DNA in Maternal Plasma (with Supporting Information)", PNAS, vol. 105, No. 51, 2008, 20458-20463.

PNAS, vol. 105, No. 51, 2008, 20458-20463. Choi, M. et al., "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing", PNAS, vol. 106, No. 45, Nov. 10, 2009, 19096-19101.

Choi, Y. et al., "Comparison of phasing strategies for whole human genomes", PLOS Genetics, Apr. 5, 2018, 26 pages.
Chothia, C. et al., "Canonical structures for the hypervariable

Chothia, C. et al., "Canonical structures for the hypervariable regions of immunoglobulins", Journal of Molecular Biology, vol. 196, No. 4, Aug. 20, 1987, 901-917.

Chu, T. et al., "Statistical Considerations for Digital Approaches to Non-Invasive Fetal Genotyping", Bioinformatics (Advance Access publication), 26 (22), 2010, 2863-2866.

Chu, Tianjiao et al., "Statistical Model for Whole Genome Sequencing and its Application to Minimally Invasive Diagnosis of Fetal Genetic Disease", Bioinformatics, 25(10), 2009, 1244-1250.

Chu, Tianjiao. et al., "A Novel Approach Toward the Challenge of Accurately Quantifying Fetal DNA in Maternal Plasma", Prenatal Diagnosis, 30, 2010, 1226-1229.

Chun, et al., "Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene", Nucleic Acids Research, vol. 35, No. 6, 2007, 1-6.

Chung, G. T. et al., "Lack of Dramatic Enrichment of Fetal DNA in Maternal Plasma by Formaldehyde Treatment", Clinical Chemistry, vol. 51, No. 2, 2005, 655, 658.

vol. 51, No. 3, 2005, 655-658. Church, et al., "Multiplex DNA Sequencing", Science, vol. 240, No. 4849, Apr. 8, 1988, 185-188.

# US 11,530,454 B2

Page 11

### (56) References Cited

#### OTHER PUBLICATIONS

Ciriello, G. et al., "Emerging landscape of oncogenic signatures across human cancers", Nature Genetics, vol. 45, No. 10, Oct. 2013, 1127-1135

Clausen, F. B. et al., "Improvement in fetal DNA extraction from maternal plasma. Evaluation of the NucliSens Magnetic Extraction system and the QIAamp DSP Virus Kit in comparison with the QIAamp DNA Blood Mini Kit", Prenatal Diagnosis, vol. 27, 2007, 6-10.

Cole, Neal W. et al., "Hyperglycemia-Induced Membrane Lipid Peroxidation and Elevated Homocysteine Levels Are Poorly Attenuated by Exogenous Folate in Embryonic Chick Brains", Comparative Biochemistry and Physiology, Part B, 150, 2008, 338-343.

Colella, S. et al., "QuantiSNP: an Objectives Bayes Hidden-Markov Model to Detect and Accurately Map Copy Number Variation Using SNP Genotyping Data", Nucleic Acids Research, 35 (6), 2007, 2013-2025.

Conlin, L. K. et al., "Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis", Human Molecular Genetics, vol. 19, No. 7, Jan. 6, 2010, 1263-1275.

Coombes, R. C., "Abstract P4-01-02: Early detection of residual breast cancer through a robust, scalable and personalized analysis of circulating tumour DNA (ctDNA) antedates overt metastatic recurrence", Cancer Research, vol. 79, No. 4 Supplement, Feb. 15, 2019. Cossu, Gianfranco et al., "Rh D/d Genotyping by Quantitative Polymerase Chain Reaction and Capillary Zone Electrophoresis", Electrophoresis, 17, 1996, 1911-1915.

Costa, J.-M. et al., "Fetal RHD genotyping in maternal serum during the first trimester of pregnancy", British Journal of Haematology, vol. 119, 2002, 255-260.

Couraud, S. et al., "Noninvasive Diagnosis of Actionable Mutations by Deep Sequencing of Circulating Free DNA in lung Cancer from Never-Smokers: A Proof-of-Concept Study from BioCAST / IFCT-1002", Clinical Cancer Research, vol. 20, No. 17, Jul. 10, 2014, 4613-4624.

Couraud, S. et al., "Supplementary Data for Noninvasive Diagnosis of Actionable Mutations by Deep Sequencing of Circulating Free DNA in Tung Cancer from Never-Smokers: A Proof-of-Concept Study from BioCAST / IFCT-1002", 2014, 13 pages.

Coyle, J. F. et al., "Standards for detailed clinical models as the basis for medical data exchange and decision support", International Journal of Medical Informatics, 69(2-3), 2003, 157-174.

Craig, D. W. et al., "Identification of genetic variants using barcoded multiplexed sequencing", Nature Methods, vol. 5, Oct. 2008, 887-893.

Crespo-Leiro, et al., "Gene Expression Profiling for Monitoring Graft Rejection in Heart Transplant Recipients", Transplantation Proceedings vol 41 No. 6 Jul 1, 2009, 2240-2243

Proceedings, vol. 41, No. 6, Jul. 1, 2009, 2240-2243.
Croft, Jr., Daniel et al., "Performance of Whole-Genome Amplified DNA Isolated from Serum and Plasma on High-Density Single Nucleotide Polymorphism Arrays", Journal of Molecular Diagnostics, 10(3), 2008, 249-257.

Cronn, R. et al., "Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology", Nucleic Acids Research, vol. 36, No. 19, Aug. 27, 2008, 11 pgs.

Cross, Jillian et al., "Resolution of trisomic mosaicism in prenatal diagnosis: estimated performance of a 50K SNP microarray", Prenat Diagn 2007; 27, 2007, 1197-1204.

Cunningham, K. S. et al., "An approach to endomyocardial biopsy interpretation", Journal of Clinical Pathology, vol. 59, No. 2, Mar. 2006, 121-129

Dahl, et al., "Multigene Amplification and Massively Parallel Sequencing for Cancer Mutation Discovery", Proceedings of the National Academy of Sciences, vol. 104, No. 22, May 29, 2007, 9387-9392. Dambrin, et al., "A New Rejection Criteria in the Heterotopically Placed Rat Heart by Non-invasive Measurement oOf Dp/Dtmax", The Journal of Heart and Lung Transplantation, vol. 18, No. 6, Jun. 18, 1999, 524-531.

Daniels, G. et al., "Fetal blood group genotyping from DNA from maternal plasma: an important advance in the management and prevention of haemolytic disease of the fetus and newborn", Vox Sanguinis, vol. 87, 2004, 223-232.

D'Aquila, Richard et al., "Maximizing sensitivity and specificity of PCR by pre-amplification heating", Nucleic Acids Research, 19(13), 1991, p. 3749.

Daruwala, Raoul-Sam et al., "A Versatile Statistical Analysis Algorithm to Detect Genome Copy Number Variation", PNAS, 101(46), 2004, 16292-16297.

Dawson, S.J. et al., "Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer", The New England Journal of Medicine, vol. 368, No. 13, Mar. 28, 2013, 1199-1209.

De Bruin, E. et al., "Spatial and temporal diversity in genomic instability processes defines lung cancer evolution", Science, vol. 346, No. 6206, Oct. 10, 2014, 251-256.

346, No. 6206, Oct. 10, 2014, 251-256. De Jong, M. M. et al., "Genes other than BRCA 1 and BRCA2 involved in breast cancer susceptibility", J. Med. Genet., vol. 39, 2009, 225-242.

De Vries, et al., "Diagnostic genome profiling in mental retardation", Am J Hum Genet, 77, published online Aug. 30, 2005, 2005, 606-616.

Deangelis, M. et al., "Solid-phase Reversible Immobilization for the Isolation of PCR Products", Nucleic Acids Research, 23 (22), 1995, 4742-4743.

Deb, Mahua et al., "Development of a Multiplexed PCR Detection Method for Barley and Cereal Yellow Dwarf Viruses, Wheat Spindle Streak Virus, Wheat Streak Mosaic Virus and Soil-Borne Wheat Mosaic Virus", Journal of Virological Methods, vol. 148, 2008, pp. 17-24.

Delaneau, O. et al., "Shape-IT: new rapid and accurate algorithm for haplotype inference", BMC Bioinformatics, vol. 9, No. 540, Dec. 16, 2008, 14 pages.

Delgado, P. O. et al., "Characterization of cell-free circulating DNA in plasma in patients with prostate cancer", Tumor Biol., vol. 34, 983-986, 2013.

Deng, S. et al., "TNER: A Novel Background Error Suppression Method for Mutation Detection in Circulating Tumor DNA", bioRxiv, http://dx.doi.org/10.1101/214379, Nov. 5, 2017, 12 pgs. Deusen, et al., "Comprehensive Detection of Driver Mutations in

Deusen, et al., "Comprehensive Detection of Driver Mutations in Acute Myeloid Leukemia Including Internal Tandem Duplications with Anchored Multiplex PCR and Next-Generation Sequencing", Blood, vol. 128, No. 22, 2016, 5251.

Blood, vol. 128, No. 22, 2016, 5251.

Deutsch, S. et al., "Detection of aneuploidies by paralogous sequence quantification", J Med Genet, vol. 41, 2004, 908-915.

Devaney, S. et al., "Noninvasive Fetal Sex Determination Using

Devaney, S. et al., "Noninvasive Fetal Sex Determination Using Cell-Free Fetal DNA: A Systematic Review and Meta-analysis", JAMA, 306(6), 2011, 627-636.

Dhallan, et al., "Methods to Increase the Percentage of Free Fetal DNA Recovered from the Maternal Circulation", JAMA, 291(9), 2004. 1114-1119.

Dhallan, Ravinder et al., "A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study", The Lancet, 369, 2007, 474-481.

Di, X. et al., "Dynamic model based algorithms for screening and genotyping", Bioinformatics, vol. 21, No. 9, 2005, 1958-1963.

Dias-Santagata, D. et al., "BRAF V600E Mutations Are Common in Pleomorphic Xanthoastrocytoma: Diagnostic and Therapeutic Implications", PLoS One, vol. 6, No. 3, Mar. 2011, 9 pages.

Diaz, et al., "Liquid Biopsies: Genotyping Circulating Tumor DNA", Journal of Clinical Oncology, vol. 32, No. 6, 2014, 579-586.

Dickover, R. E. et al., "Optimization of Specimen-Handling Procedures for Accurate Quantitation of Levels of Human Immunodeficiency Virus RNA in Plasma by Reverse Transcriptase PCR", Journal of Clinical Microbiology, vol. 36, No. 4, 1998, 1070-1073, Dieffenbach, C W. et al., "General concepts for PCR primer design", Genome Research. PCR methods and Applications vol. 3, 1993, S30-S37.

Diehl, F. et al., "Circulating mutant DNA to assess tumor dynamics", Nature Medicine, vol. 14, No. 9, Jul. 31, 2008, 985-990. Diehl, F. et al., "Detection and quantification of mutations in the plasma of patients with colorectal tumors", PNAS, vol. 102, No. 45, Nov. 8, 2005, 16368-16373.

Case: 24-1324 Document: 42-1 Page: 326 Filed: 03/18/2024

# US 11,530,454 B2

Page 12

#### References Cited (56)

#### OTHER PUBLICATIONS

Dietmaier, W. et al., "Multiple Mutation Analyses in Single Tumor Cells with Improved Whole Genome Amplification", American Journal of Pathology, vol. 154, No. 1, Jan. 1999, 83-95

Ding, C. et al., "Direct molecular haplotyping of long-range genomic DNA with M1-PCR", PNAS 100(13), 2003, 7449-7453

Ding, C. et al., "MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diag-

nosis", PNAS, vol. 101, No. 29, Jul. 20, 2004, 10762-10767. Dodge, Y., "Bayes' Theorem", The Concise Encyclopedia of Statistics, 2008, 30-31.

Dohm, J. et al., "Substantial Biases in Ultra-Short Read Data Sets From High-Throughput DNA Sequencing", Nucleic Acids Research, 36 (16), e105, 2008, 10 pgs.

Dolganov, Gregory et al., "A Novel Method of Gene Transcript Profiling in Airway Biopsy Homogenates Reveals Increased Expression of a Na-K+-Cl-Cotransporter (NKCC1) in Asthmatic Subjects", Genome Res., 11, 2001, 1473-1483.

Donaghue, C. et al., "Detection of mosaicism for primary trisomies in prenatal samples by QF-PCR and karyotype analysis", Prenatal Diagnosis, vol. 25, 2005, 65-72.

Donohoe, Gerard G. et al., "Rapid Single-Tube Screening of the C282Y Hemochromatosis Mutation by Real-Time Multiplex Allelespecific PCR without Fluorescent Probes", Clinical Chemistry, 46, 10, 2000, 1540-1547.

Donoso, P. et al., "Current Value of Preimplantation Genetic Aneuploidy Screening in IVF", Human Reproduction Update, 13(1), 2007,

Doostzadeh, et al., "High Throughput Automated Allele Frequency Estimation by Pyrosequencing", PLoS ONE, vol. 3, No. 7, Jul. 16, 2008, 1-4.

Dorit, D. L., "cDNA Amplification Using One-sided (Anchored) Pcr", Current Protocols in Molecular Biology, vol. 17, 1992, pp.

Dorit, Robert L. et al., "One-sided Anchored Polymerase Chain Reaction for Amplification and Sequencing of Complementary DNA", Methods in Enzymology, vol. 218 1993, pp. 36-47.

Dowd, P. et al., "On the mechanism of the anticlotting action of vitamin R quinone", Proc. Natl. Acad. Sci. USA, vol. 92, 1995, 8171-8175.

Downward, J., "Targeting Ras Signalling Pathways in Cancer Therapy", Nature Reviews, vol. 3, Jan. 2003, 11-22. Dressman, D. et al., "Transforming single DNA molecules into

fluorescent magnetic particles for detection and enumeration of genetic variations", PNAS, vol. 100, No. 15, Jul. 22, 2003, 8817-8822

Echeverri, et al., "Caffeine's Vascular Mechanisms of Action", International Journal of Vascular Medicine vol. 2010(2010), 10 pages, Aug. 25, 2010.

pages, Aug. 25, 2010. Edwards, M. C. et al., "Multiplex PCR: Advantages, Development, and Applications", Genome Research, vol. 3, 1994, S65-S75. Efron, B. et al., "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy", Statistical Science, vol. 1, No. 1, 1986, 54-77. Ehlayel, A. et al., "Emerging monitoring technologies in kidney

transplantation", Pediatric Nephrology, vol. 36, 2021, 3077-3087. Ehrich, Mathias et al., "Noninvasive Detection of Fetal Trisomy 21 by Sequencing of DNA in Maternal Blood: A Study in a Clinical Setting", American Journal of Obstetrics & Gynecology, 204, 2011, 205.e1-205.811.

Eichler, H, "Mild Course of Fetal Rh D Haemolytic Disease due to Maternal Alloimmunisation to Paternal HLA Class I and II Antigens", Vox Sang, 68, 1995, 243-247.

Ellison, Aaron M., "Bayesian Inference in Ecology", Ecology Letters, vol. 7, 2004, 509-520.

Ellonen, P. et al., "Development of SNP Microarray for Supplementary Paternity Testing", International Congress Series, 1261, 2004, 12-14

Elnifro, Elfath M., "Multiplex PCR: Optimization and Application in Diagnostic Virology", Clinical Microbiology Reviews, vol. 13, 2000, pp. 559-570.

Eltoukhy, H. et al., "Modeling and Base-Calling for DNA Sequencing-By-Synthesis", IEEE, 2006, II-1032-II-1035

EP06838311.6, "European Communication and Extended European Search Report", dated Dec. 30, 2008, 8 pgs.

EP08742125.1, "European Communication pursuant to Article 94(3) EPC and Examination Report", dated Feb. 12, 2010, 5 pgs.

Erijman, Ariel et al., "Transfer-PCR (TPCR): A Highway for DNA Cloning and Protein Engineering", Journal of Structural Biology,

vol. 175, 2011, pp. 171-177. Erlich, R. L. et al., "Next-generation sequencing for HLA typing of class loci", BMC Genomics, vol. 12, No. 42, 2011, 13 pages

Eronen, L. et al., "HaploRec: efficient and accurate large-scale reconstruction of haplotypes", BMC Bioinformatics, vol. 7, No.

542, Dec. 22, 2006, 18 pages. European Commission, "The 7th International Conference on Circulating Nucleic Acids in Plasma and Serum (CNAPS VII) in Madrid—Spain", The International Conference on Circulating Nucleic Acids in Plasma and Serum, Oct. 24, 2011, 2 pgs. Everitt, B. S. "Medical Statistics From A to Z", 2003, 3 pages.

Fackenthal, J. D. et al., "Aberrant RNA splicing and its functional consequences in cancer cells", Disease Models & Mechanisms, vol. 1, 2008, 37-42.

Faham, M. et al., "Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia", Blood Journal, vol. 120, No. 26, Dec. 20, 2012, 5173-5180.

Falcon, O., "Screening for trisomy 21 by fetal tricuspid regurgitation, nuchal translucency and maternal serum free b-hCG and PAPP-A at 11+0 to 13+6 weeks", Ultrasound Obstet Gynecol, vol. 27, 2006, 151-155.

Fan, et al., "Whole-genome molecular haplotyping of single cells", Nature Biotechnology, vol. 29, No. 1, Jan. 1, 2011, 51-5

Fan, C H. et al., "Detection of Aneuploidy with Digital Polymerase Chain Reaction", Analytical Chemistry, vol. 79, No. 19, Oct. 1, 2007, 7576-7579.

Fan, Christina H. et al., "Non-Invasive Prenatal Measurement of the Fetal Genome", Nature, doi:10.1038/nature11251, 2012, 26 pgs. Fan, Christina H et al., "Noninvasive Diagnosis of Fetal Aneuploidy by Shotgun Sequencing DNA from Maternal Blood", PNAS, 105, 42, 2008, 16266-16271.

Fan, H. C. et al., "In Principle Method for Noninvasive Determination of the Fetal Genome", Nat. Prec., 2010, 16 pgs.

Fan, H. C. et al., "Microfluidic digital PCR enables rapid prenatal diagnosis of fetal aneuploidy", American Journal of Obstetrics & Gynecology, vol. 200, May 2009, 543.e1-543.e7.
Fan, H. Christina et al., "Sensitivity of Noninvasive Prenatal

Detection of Fetal Aneuploidy from Maternal Plasma Using Shotgun Sequencing Is Limited Only by Counting Statistics", PLoS ONE, vol. 5, Issue 5 (e10439), May 3, 2010, 1-6. Fan, J.-B. et al., "Highly Parallel SNP Genotyping", Cold Spring Harbor Symposia on Quantitative Biology, vol. LXVIII, Feb. 2003,

Fan, Jian-Bing et al., "Highly Parallel Genomic Assay", Nature Reviews, 7, 2006, 632-644.

Fat Secret, "5 Foods to Never Eat", https://www.fatsecret.com/ calories-nutrition/food/white-bread/carboyhydrate (printed from internet Nov. 1, 2014)., 2 pages. Fazio, Gennaro, et al., "Identification of RAPD Markers Linked to

Fusarium Crown and Root Rot Resistance (Frl) in Tomato", Euphytica 105, 1999, 205-210.

Findlay, I. et al., "Allelic drop-out and preferential amplification in single cells and human blastomeres: implications for preimplantation diagnosis of sex and cystic fibrosis", Molecular Human Reproduction, vol. 1, 1995, 1609-1618. Fiorentino, F. et al., "Development and Clinical Application of a

Strategy for Preimplantation Genetic Diagnosis of Single Gene Disorders Combined with HLA Matching", Molecular Human Reproduction (Advance Access publication), 10 (6), 2004, 445-460. Fiorentino, F et al., "Strategies and Clinical Outcome of 250 Cycles of Preimplantation Genetic Diagnosis for Single Gene Disorders", Human Reproduction, 21, 3, 2006, 670-684.

Document: 42-1 Case: 24-1324 Page: 327 Filed: 03/18/2024

# US 11,530,454 B2

Page 13

#### References Cited (56)

#### OTHER PUBLICATIONS

Fiorentino, Francesco et al., "Short Tandem Repeats Haplotyping of the HLA Region in Preimplantation HLA Matching", European Journal of Human Genetics, 13, 2005, 953-958.

Fitzgerald, "Intravascular Ultrasound Imaging of Coronary Arteries is Three Layers the Norm?", Circulation, vol. 86, No. 1, Jul. 1, 1992,

Ford, E. et al., "A method for generating highly multiplexed ChIP-seq Tibraries", BMC Research Notes, vol. 7, No. 312, May 22, 2014, 1-5.

Forejt, et al., "Segmental trisomy of mouse chromosome 17: introducing an alternative model of Down's syndrome", Genomics, 4(6),

Forshew, et al., "Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA", Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci. Transl. Med. 4, 136 30 (2012)., 1-12.

Forshew, T. et al., "Supplementary Materials for Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA", Sci. Transl. Med, vol. 4, May 30, 2012, 20 pgs

Fortina, P. et al., "Detection of the most common mutations causing beta-thalassemia in Mediterraneans using a multiplex amplification Refractory mutation system (MARMS)", Genome Res., vol. 2, 1992, 163-166.

Fortina, P. et al., "DOP-PCR Amplification of Whole Genomic DNA and Microchip-Based Capillary Electrophoresis", Methods in Molecular Biology: Capillary Electrophoresis of Nucleic Acids, vol. II Practical Applications of Capillary Electrophoresis, 2001, 211-

Fouquet, C. et al., "Rapid and Sensitive p53 Alteration Analysis in Biopsies from Lung Cancer Patients Using a Functional Assay and a Universal Oligonudeotide Array: A Prospective Study", Clinical Cancer Research, vol. 10, May 15, 2004, 3479-3489.

Fournie, et al., "Plasma DNA as a Marker of Cancerous Cell Death. Investigations in Patients Suffering From Lung Cancer and in Nude Mice Bearing Human Tumours", Cancer Letters, vol. 91, No. 2, May 8, 1995, 221-227

Fredriksson, et al., "Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector", Nucleic Acids Research, 2007, vol. 35, No. 7e47, 1-6. Fredriksson, M. et al., "Assessing Hematopoietic Chimerism After

Allogeneic Stem Cell Transplantation by Multiplexed SNP Genotyping Using Microarrays and Quantitive Analysis of SNP Alleles", Leukemia, vol. 18, Issue 2, Dec. 4, 2003, 255-266.

Freeman, Jennifer L. et al., "Copy Number Variation: New Insights in Genome Diversity", Genome Research, 16, 2006, 949-961. Frohman, M.A. et al., "On Beyond Classic RACE (Rapid Ampliance)".

fication of cDNA Ends)", Genome Research, vol. 4, 1994, S40-S58. Frost, Mackenzie S et al., "Differential Effects of Chronic Pulsatile Versus Chronic Constant Maternal Hyperglycemia on Fetal Pancreatic B-Cells", Journal of Pregnancy, 2012,, Article ID 812094, 2012,

Fu, G. K. et al., "Counting individual DNA molecules by the stochastic attachment of diverse labels", PNAS, vol. 108, No. 22, May 31, 2011, 9026-9031.

Fu, G. K. et al., "Digital Encoding of Cellular mRNAs Enabling Precise and Absolute Gene Expression Measurement by Single-Molecule Counting", Analytical Chemistry, vol. 86, Mar. 3, 2014,

Fu, Yao-Wen et al., "Presence of Donor-and-recipientderived Dna Microchimerism in the Cell-free Blood Samples of Renal Transplantation Recipients Associates With the Acceptance of Transplanted Kidneys", Asian Journal of Andrology, vol. 8, No. 4, Jul. 1, 2006, 477-482.

Gadi, V. K. et al., "Soluble Donor DNA Concentrations in Recipient Serum Correlate with Pancreas-Kidney Rejection", Clinical Chemistry, vol. 52, No. 3, 2006, 379-382.

Ganshirt-Ahlert, D. et al., "Ratio of Fetal to Maternal DNA is Less Than 1 in 5000 at different Gestational Ages in Maternal Blood", Clinical Genetics, 38, 1990, 38-43.

Ganshirt-Ahlert, D. et al., "Fetal DNA in Uterine Vein Blood",

Obstetrics & Gynecology, 80 (4), 1992, 601-603. Ganshirt-Ahlert, Dorothee et al., "Three Cases of 45,X/46,XYnf Mosaicism", Human Genetics, 76, 1987, 153-156.

Gao, et al., "Relation of Donor Age and Preexisting Coronary Artery Disease on Angiography and Intracoronary Ultrasound to Later Development of Accelerated Allograft Coronary Artery Disease", The American Journal of Cardiology, vol. 29, No. 3, Mar. 1, 1997, 623-629.

Gao, F. et al., "Characterizing Immunoglobulin Repertoire from Whole Blood by a Personal Genome Sequencer", PLOS One, vol. 8, No. 9, Sep. 13, 2013, 8 pgs.

Gao, Ming et al., "Characterization of dull1, a Maize Gene Coding for a Novel Starch Synthase", The Plant Cell, vol. 10, 1998, pp.

Garcia Moreira, V. et al., "Cell-Free DNA as a Noninvasive Acute Rejection Marker in Renal Transplantation", Clinical Chemistry, vol. 55, No. 11,2009, 1958-1966.

Garcia-Murillas, I. et al., "Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer", Science Translational Medicine, vol. 7, No. 302, Aug. 26, 2015, 34 pgs.

Gardina, P. et al., "Ploidy Status and Copy Number Aberrations in Primary Glioblastomas Defined by Integrated Analysis of Allelic Ratios, Signal Ratios and Loss of Heterozygosity Using 500K SNP Mapping Arrays", BMC Genomics, 9 (489), (doi:10.1186/1471-2164-9-489), 2008, 16 pgs.

Gautier, E. et al., "Fetal RhD genotyping by maternal serum analysis: A two-year experience", American Journal of Obstetrics and Gynecology, vol. 192,2005, 666-669.

Geifman-Holtzman, et al., "Prenatal Diagnosis: Update on Invasive Versus Noninvasive Fetal Diagnostic Testing From Maternal Blood", Expert Review of Molecular Diagnostics, vol. 8, No. 6, Nov. 1, 2008, 727-751.

Geiss, G. K. et al., "Direct multiplexed measurement of gene expression with color-coded probe pairs", Nature Biotechnology, vol. 26, No. 3, Feb. 17, 2008, 317-325.

Ghanta, Sujana et al., "Non-Invasive Prenatal Detection of Trisomy 21 Using Tandem Single Nucleotide Polymorphisms", PLoS ONE, 5 (10), 2010, 10 pgs

Gholami, M. et al., "A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants", Plant Biotechnology Journal, vol. 10, 2012, 635-645.

Gielis, E. M. et al., "Cell-Free DNA: An Upcoming Biomarker in Transplantation", American Journal of Transplantation, vol. 15, 2015, 2541-2551.

Gielis, E. M. et al., "Plasma donor-derived cell-free DNA kinetics after kidney transplantation using a single tube multiplex PCR assay", PLOS One, vol. 13, No. 12, e0208207, Dec. 6, 2018, 16 pgs. Gineikiene, Egle et al., "Single Nucleotide Polymorphism-based System Improves The Applicability of Quantitative PCR for Chimerism Monitoring", Journal of Molecular Diagnostics, vol. 11, No. 1, Jan. 1, 2009, 66-74.

Gingeras, et al., "Fifty Years of Molecular (DNA/RNA) Diagnostics", Clinical Chemistry, vol. 51, No. 3, Jan. 13, 2005, 661-671. Girnita, Diana M. et al., "Disparate Distribution of 16 Candidate Single Nucleotide Polymorphisms Among Racial and Ethnic Groups of Pediatric Heart Transplant Patients", Transplantation, vol. 82, No. 12, Dec. 27, 2006, 1774-1780.

Gjertson, David W. et al., "Assessing Probability of Paternity and the Product Rule in DNA Systems", Genetica, 96, 1995, 89-98.

Glaab, W. E. et al., "A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan) and mismatch amplification mutation assay", Mutation Research, vol. 430, 1999, 12 pgs.

Gnirke, A. et al., "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing", Nature Biotechnology, vol. 27, No. 2, Feb. 2009, 182-189.

# US 11,530,454 B2

Page 14

### (56) References Cited

#### OTHER PUBLICATIONS

Go, A. T. et al., "Non-invasive aneuploidy detection using free fetal DNA and RNA in maternal plasma: recent progress and future possibilities", Human Reproduction Update, vol. 17, No. 3, 2011, 372-382.

Goncalves-Primo, A. et al., "Investigation of Apoptosis-Related Gene Expression Levels in Preimplantation Biopsies as Predictors of Delayed Kidney Graft Function", Transplantation, vol. 97, No. 12, Jun. 27, 2014

12, Jun. 27, 2014. Gordon, et al., "Disease-Specific Motifs Can Be Identified in Circulating Nucleic Acids From Live Elk and Cattle Infected With Transmissible Spongiform Encephalopathies", Nucleic Acids Research, vol. 37. No. 2, Feb. 1, 2009, 550-556.

Gorringe, et al., "Large-scale Genomic Analysis of Ovarian Carcinomas", Molecular oncology, vol. 3, No. 2, Apr. 1, 2009, 157-164. Gouya, et al., "Coronary Artery Stenosis In High-risk Patients: 64-section Ct and Coronary Angiography—Prospective Study and Analysis of Discordance", Radiology, vol. 252, No. 2, Aug. 1, 2009, 377-385.

Greenwalt, T. et al., "The Quantification of Fetomaternal Hemorrhage by an Enzyme-Linked Antibody Test with Glutaraldehyde Fixation", Vox Sang, 63, 1992, 268-271.

Gregory, et al., "Comparison of Sixty-Four-Slice Multidetector Computed Tomographic Coronary Sngiography to Coronary Angiography With Intravascular Ultrasound for the Detection of Transplant Vasculopathy", The American Journal of Cardiology, vol. 98, No. 7, Aug. 4, 2006, 877-884.

Grenda, R. "Torque teno (TTV) viral load as a biomarker of

Grenda, R. "Torque teno (TTV) viral load as a biomarker of immunosuppressive strength after kidney transplantation in children", Pediatric Nephrology, vol. 36, May 27, 2020, 3 pages.

Griffiths, A. J. et al., "An Introduction to Genetic Analysis", Sixth Edition, 1996, 5 pages.
Grskovic, M. et al., "Validation of a Clinical-Grade Assay to

Grskovic, M. et al., "Validation of a Clinical-Grade Assay to Measure Donor-Derived Cell-Free DNA in Solid Organ Transplant Recipients", The Journal of Molecular Diagnostics, vol. 18, No. 6 + Supplemental Appendix S1, Nov. 2016, 890-902.

Grunenwald, H., "Optimization of Polymerase Chain Reactions", Methods in Biology, vol. 226, 2003, 89-99.

Gu, H. et al., "Diagnostic role of microRNA expression profile in the serum of pregnant women with fetuses with neural tube defects", Journal of Neurochemistry, vol. 122, 2012, 641-649.

Guerra, J., "Terminal Contributions for Duplex Oligonucleotide Thermodynamic Properties in the Context of Nearest Neighbor Models" Biopolymers 95(3) (2010) 2011 194-201

Models", Biopolymers, 95(3), (2010), 2011, 194-201. Guetta, Esther et al., "Analysis of Fetal Blood Cells in the Maternal Circulation: Challenges, Ongoing Efforts, and Potential Solutions", Stem Cells and Development, 13, 2004, 93-99.

Guichoux, et al., "Current Trends in Microsatellite Genotyping", Molecular Ecology Resources, 11, 2011, 591-911.

Gunderson, K. L. et al., "A genome-wide scalable SNP genotyping assay using microarray technology", Nature Genetics, vol. 37, No. 5, May 2005, 549-554.

Gundry, C. N. et al., "Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplicons", Nucleic Acids Research, vol. 36, No. 10, Apr. 29, 2008, 3401-3408. Guo, H et al., "A Specific and Versatile Genome Walking Technique", Gene, vol. 381, 2006, 18-23.

Gusella, J. et al., "Precise localization of human B-globin gene complex on chromosome 11\*", Proc. Natl. Acad. Sci USA, vol. 76, No. 10, Oct. 1979, 5239-5243.

Gwee, Pai-Chung et al., "Simultaneous Genotyping of Seven Single-nucleotide Polymorphisms in the Mdr1 Gene by Single-tube Multiplex Minisequencing", Pai-Chung Gwee. et al., "Simultaneous Genotyping of Seven Single-nucleotide Polymorphisms in the Mdr1 Gene by Single-tube Multiplex Minisequencing", Clinical chemistry, Apr. 2003, vol. 49, Issue. 3, pp. 672-676, Apr. 1, 2003, 672-676. Hahn, et al., "Non-invasive Prenatal Diagnostics Using Next Generation Sequencing: Technical, Legal And Social Challenges", Expert Opinion on Medical Diagnostics, vol. 6, No. 6, Nov. 1, 2012, 517-528.

Hahn, S. et al., "Current applications of single-cell PCR", CMLS Cellular and Molecular. Life Sciences, vol. 57, 2000, 96-105.

Hahn, S. et al., "Quantification of Circulating DNA: In the Preparation Lies the Rub", Clinical Chemistry, vol. 47, No. 9, 2001, 1577-1578.

Hainer & Fazzio, "High-Resolution Chromatin Profiling Using CUT&RUN", Current Protocols in Molecular Biology, 2019, 1-22. Halford, William P., "The Essential Prerequisites for Quantitative RT-PCR", Nature Biotechnology, vol. 17, 1999, 1 page.

Hall, M., "Panorama Non-Invasive Prenatal Screening for Microdeletion Syndromes", Apr. 1, 2014 (Apr. 1, 2014), XP055157224, Retrieved from the Internet: URL:http://www.panoramatest.com/sites/default/files/files/PanoramaMicrodeletionsWhite Paper-2.pdf Iretrieved on Dec. 8, 2014].

[retrieved on Dec. 8, 2014]. Han, S-W et al., "Predictive and Prognostic Impact of Epidermal Growth Factor Receptor Mutation in Non-Small-Cell Lung Cancer Patients Treated With Gefitinib", Journal of Clinical Oncology, vol. 23, No. 11, Apr. 10, 2005, 2493-2501.

Handley, D. et al., "Noninvasive prenatal chromosomal aneuploidy detection using plasma cell-free nucleic acid", Expert Rev Obstet. Gynecol, vol. 5, No. 5, 2010, 581-590. Handyside, et al., "Isothermal whole genome amplification from

Handyside, et al., "Isothermal whole genome amplification from single and small Nos. of cells: a new era for preimplantation genetic diagnosis of inherited disease", Molecular Human Reproduction vol. IO, No. 10 pp. 767-772, 2004.

Hao, T. B. et al., "Circulating cell-free DNA in serum as a biomarker

Hao, T. B. et al., "Circulating cell-free DNA in serum as a biomarker for diagnosis and prognostic prediction of colorectal cancer", British Journal of Cancer, vol. 111, Aug. 26, 2014, 1482-1489.

Hara, Eiji et al., "Subtractive eDNA cloning using oligo(dT)3o-latex and PCR: isolation of eDNA clones specific to undifferentiated human embryonal carcinoma cells", Nucleic Acids Research, 19(25), 1991, 7097-7104.

Hardenbol, P., "Multiplexed Genotyping With Sequence-Tagged Molecular Inversion Probes", Nature Biotechnology, 21 (6), 2003, 673-678.

Hardenbol, Paul et al., "Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a singled tube assay", Genome Research, 15, 2005, 269-275.

Harismendy, O. et al., "Method for Improving Sequence Coverage Uniformity of Targeted Genomic Intervals Amplified by LR-PCR Using Illumina GA Sequencing-By-Synthesis Technology", Bio Techniques, 46(3), 2009, 229-231.

Techniques, 46(3), 2009, 229-231.

Harper, J. C. et al., "Recent Advances and Future Developments in PGD". Prenatal Diagnosis, 19, 1999, 1193-1199.

PGD", Prenatal Diagnosis, 19, 1999, 1193-1199. Harton, G.L. et al., "Preimplantation Genetic Testing for Marfan Syndrome", Molecular Human Reproduction, 2 (9), 1996, 713-715. Hartwell, L. H. et al., "Chapter 11: The Direct Detection of Genotype Distinguishes Individual Genomes", Genetics: From Genes to Genomes, Second Edition, 2004, 371-414.

Hartwell, L. H. et al., "Chapter 13: Chromosomal Rearrangements and Changes in Chromosome Number Reshape Eukaryotic Genomes", Genetics: From Genes to Genomes, Second Edition, 2004, 441-486. Hattori, M. et al., "The DNA sequence of human chromosome 21", Nature, vol. 405, May 18, 2000, 311-319.

Hawkins, T. et al., "Whole genome amplification—applications and advances", Current Opinion in Biotechnology, 13, 2002, 65-67. Hayden, et al., "Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping", BMC Genomics 2008, 9(80),

He, QZ et al., "A method for improving the accuracy of non-

invasive prenatal screening by cell-free foetal DNA size selection", British Journal of Biomedical science, vol. 75, No. 3, Jul. 2018, 133-138.

Heaton, Paul R. et al., "Heminested PCR Assay for Detection of Six Genotypes of Rabies and Rabies-related Viruses", Journal of Clinical Microbiology, vol. 35, 1997, pp. 2762-2766.

Heidary, M. et al., "The dynamic range of circulating tumor DNA in metastatic breast cancer", Breast Cancer Research, vol. 16, No. 421, 2014, 10 pages.

Hellani, A. et al., "Clinical Application of Multiple Displacement Amplification in Preimplantation Genetic Diagnosis", Reproductive BioMedicine Online, 10 (3), 2005, 376-380.

Case: 24-1324 Document: 42-1 Page: 329 Filed: 03/18/2024

# US 11,530,454 B2

Page 15

#### References Cited (56)

#### OTHER PUBLICATIONS

Hellani, Ali et al., "Multiple displacement amplification on single cell and possible PGD applications", Molecular Human Reproduction, 10(11), 2004, 847-852

Henegariu, O. et al., "Multiplex PCR: Critical Parameters and Step-by-Step Protocol", Biotechniques, vol. 23, 1997, pp. 504-511. Hidestrand, M. et al., "Highly Sensitive Noninvasive Cardiac Transplant Rejection Monitoring Using Targeted Quantification of Donor-Specific Cell-Free Deoxyribonucleic Acid", Journal of the American College of Cardiology, vol. 63, No. 12, 1224-1226, 2014.

Hiendleder, et al., "Functional genomics: tools for improving farm animal health and welfare", Rev. Sci. Tech. Off. Int. Epiz., 24 (1),

Hoberman, Rose et al., "A Probabilistic Approach for SNP Discovery in High-throughput Human Resequencing Data", Genome Research, vol. 19, Jul. 15, 2009, 1542-1552.

Hochberg, et al., "A Novel Rapid Single Nucleotide Polymorphism (SNP)-Based Method for Assessment of Hematopoietic Chimerism After Allogeneic Stem Cell Transplantation", Blood, vol. 101, No. 1, Jan. 1, 2003, 363-369.

Hodges, et al., "Genome-wide In Situ Exon Capture for Selective Resequencing", Nature Genetics, vol. 39, No. 12, Nov. 4, 2007, 1522-1527.

Hodgkinson, C. L. et al., "Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer", Nature Medicine, vol. 20, No. 8, Aug. 2014, 897-905.

Hoffmann, Steven et al., "Donor Genomics Influence Graft Events the Effect of Donor Polymorphisms on Acute Rejection and Chronic Allograft Nephropathy", Kidney International, vol. 66, No. 4, Oct. 1, 2004, 1686-1693.

Hojsgaard, S et al., "BIFROST-Block recursive models induced from relevant knowledge, observations, and statistical techniques", Computational Statistics & Data Analysis, 19(2), 1995, 155-175. Hollas, B. et al., "A stochastic approach to count RN A molecules using DNA sequencing methods", Lecture Notes in Computer Science, vol. 2812, 2003, 55-62.

Holleley, et al., "Multiplex Manager 1.0: a Cross-Platform Computer Program that Plans and Optimizes Multiplex PCR", BioTechniques46:511-517 (Jun. 2009), 511-517

Hollox, E. et al., "Extensive Normal Copy Number Variation of a β-Defensin Antimicrobial-Gene Cluster", Am. J. Hum. Genet., 73, 2003, 591-600.

Holt, et al., "Detecting SNPS And Estimating Allele Frequencies in Clonal Bacterial Populations by Sequencing Pooled DNA", Bioin-

formatics, vol. 25, No. 16, Aug. 15, 2009, 2074-2075. Homer, et al., "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays", PLOS Genetics, 4(8), 2008, 9 pgs.

Hoogendoorn, Bastiaan et al., "Genotyping Single Nucleotide Polymorphisms by Primer Extension and High Performance Liquid Chromatography", Hum Genet, 104, 1999, 89-93. Horai, et al., "Novel Implantable Device to Detect Cardiac Allograft

Rejection", Circulation, vol. 120, No. Suppl 1, Sep. 15, 2009,

Hornak, M. et al., "Aneuploidy Detection in Pigs Using Comparative Genomic Hybridization: From the Oocytes to Blastocysts", PLoS ONE, vol. 7, No. 1, Jan. 2012, 6 pages. Hosmillo, Myra D. et al., "Development of Universal SYBR Green

Real-time RT-PCR for The Rapid Detection and Quantitation of Bovine and Porcine Toroviruses", Journal of Virological Methods, vol. 168, 2010, pp. 212-217

Hosono, S. et al., "Unbiased Whole-Genome Amplification Directly From Clinical Samples", Genome Research, vol. 13, 2003, 954-964. Hospital, F. et al., "A General Algorithm to Compute Multilocus Genotype Frequencies Under Various Mating Systems" vol. 12, No. 6, Jan. 1, 1996 (Jan. 1, 1996), pp. 455-462

Hou, X. et al., "Analysis of the Repertoire Features of TCR Beta Chain CDR3 in Human by High-Throughput Sequencing", Cellular Physiology and Biochemistry, vol. 39, Jul. 21, 2019, 651-667.

Howie, et al., "Fast and accurate genotype imputation in genomewide association studies through pre-phasing", Nature Genetics, vol. 44, No. 8, Jul. 22, 2012, 955-959.

Howie, B. N. et al., "A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies", PLoS Genetics, vol. 5, No. 6, Jun. 2009, 15 pages

Hu, Dong Gui et al., "Aneuploidy Detection in Single Cells Using DNA Array-Based Comparative Genomic Hybridization", Molecular Human Reproduction, 10(4), 2004, 283-289.

Hu, Hao et al., "Mutation Screening in 86 Known X-linked Mental Retardation Genes by Droplet-based Multiplex Pcr and Massive Parallel Sequencing", Hao Hu et al., "Mutation Screening in 86 Known X-linked Mental Retardation Genes by Droplet-based Multiplex Pcr and Massive Parallel Sequencing", Hugo J, Dec. 2009, vol. 3, pp. 41-49., Dec. 1, 2009, 41-49.

Hu, Y. et al., "Detection of Extrahepatic Hepatitis C Virus Replication by a Novel, Highly Sensitive, Single-Tube Nested Polymerase Chain Reaction", Am. J. Clin Pathol., vol. 119, 2003,

Huang, D. J. et al., "Reliable detection of Trisomy 21 using MALDI-TOF mass spectrometry", Genetics in Medicine, vol. 8, Nov. 2006, 728-734.

Huang, D. J. et al., "Use of an Automated Method Improves the Yield and Quality of Cell-Free Fetal DNA Extracted from Maternal Plasma", Clinical Chemistry, vol. 51, No. 12, 2005, 2419-2420.

Huang, J. et al., "Whole genome DNA copy number changes identified by high density oligonucleotide arrays", Human Genomics, vol. 1, No. 4, May 2004, 287-299.

Hubacek, et al., "Detection of Donor DNA After Heart Transplantation How Far Could It Be Affected by Blood Transfusion and Donor Chimerism?", Transplantation Proceedings, vol. 39, Jun. 1, 2007, 1593-1595.

Hug, H. et al., "Measurement of the number of molecules of a single mRNA species in a complex mRNA preparation", J. Theor. Biol., vol. 221, 2003, 615-624.

Hultin, E. et al., "Competitive enzymatic reaction to control allelespecific extensions", Nucleic Acids Research, vol. 33, No. 5, Mar. 14, 2005, 1-10.

Hung, E.C.W. et al., "Detection of circulating fetal nucleic acids: a review of methods and applications", J. Clin. PathoL, vol. 62, 2009, 308-313

Hyndman, D L. et al., "PCR Primer Design", Methods in Molecular

Biology, vol. 226, Second Edition, 2003, 81-88. Ido, Yasuo et al., "Hyperglycemia-Induced Apoptosis in Human

Umbilical Vein Endothelial Cells: Inhibition by the AMP-Activated Protein Kinase Activation", Diabetes, 51, 2002, 159-167. Illumina, "Automated GoldenGate™ Genotyping on the BeadStation 500", Pub. No. 970-2004-002, 2004, 2 pages.

Illumina, "Genomic Sequencing", Datasheet: Sequencing, 2010, 38939-38944.

Illumina, "GoldenGate" Assay Workflow: Illumina's GoldenGate assay protocol provides high-quality, high-multiplex genotyping results with a streamlined workflow, Pub. No. 370-2004-006, 2004, 2 pages.

Illumina, "HiSeq 2500 Sequencing System", System Specification Sheet: Sequencing, available via URL https://www.illumina.com/ documents/products/datasheets/datasheet\_hiseq2500.pdf, 2015, 4 pgs. Illumina, "History of Sequencing by Synthesis", https://www. illumina.com/science/technology/next-generation-sequencing/illuminasequencing-history.html, 2020, 3 pages.

Illumina, "HumanOmni1-Quad BeadChip", Illumina DNA Analysis, Pub. No. 370-21009-007, 2009, 1 page

Illumina, "HumanOmni2.5-8 BeadChips: Next-Generation GWAS Content for Genotyping and CNV Analysis", Data Sheet: DNA Analysis, Pub. No. 370-2011-008, 2011, 1 page.

Illumina, "Illumina Adapter Sequences", Published by Illumina,

Illumina, "Illumina Extends BeadArray Technology to Address Wider Range of SNP Genotyping Projects; New Microarray Offerings Enable Genotyping at 384 and 786 Multiplex", Businesswire, May 4, 2004, 2 pages

Case: 24-1324 Document: 42-1 Page: 330 Filed: 03/18/2024

# US 11,530,454 B2

Page 16

#### References Cited (56)

#### OTHER PUBLICATIONS

Illumina, "Illumina® Beadstation 500: A Scalable System That Grows With Your Research Requirements", Pub. No. 970-2005-003, Jul. 1, 2005, 4 pages

Illumina, "Illumina Announces Benchtop SNP Genotyping System", Press Release, Nov. 5, 2003, 3 pages

Illumina, "Illumina Begins Shipment of BeadStation 500G Benchtop Genotyping System", Press Release, Apr. 15, 2004, 3 pages

Illumina, "MiSeq System Information Sheet", 2018, 3 pgs. Illumina, "Patent Owner Illumina's Preliminary Response to Petition", Oct. 17, 2018, 75 pgs.

Illumina, "Petition for Inter Partes Review of U.S. Pat. No. 8,682,592",

Jun. 13, 2019, 91 pages.
Illumina, "Plaintiff/Counterclaim Defendant Illumina, Inc.'S Amended Patent L.R. 3-3 Preliminary Invalidity Contentions for U.S. Pat. No. 8,682,592", Oct. 30, 2018, 22 pages.

Illumina, "Plaintiff/Counterclaim-Defendant Illumina, Inc.'s Patent L.R. 3-3 Contentions for U.S. Patent Preliminary Invalidity Contentions for U.S. Pat. No. 8,682,592", Oct. 9, 2018, 81 pages. Illumina, "Preparing Samples for Sequencing Genomic DNA", Part # 11251892 Rev. A, 2007, 18 pages

Illumina, "Preparing Samples for Sequencing Genomic DNA", Part # 1003806 Rev. A, 2007, 20 pages.

Illumina, "Products & Services", Product Literature, Mar. 21, 2007, 3 pages.

Illumina, "Technology: Solexa Sequencing Technology", May 21,

2007, 1 page. Illumina, "TruSeq™ RNA and DNA Library Preparation Kits v2", Data Sheet: Illumina® Sequencing, 2014, 4.

Illumina Catalog, "Paired-End Sample Preparation Guide, Illumina

Catalog# PE-930-1 001, Part# 1005063 Rev. E", 2011, 1-40. Illumina, Inc., "Declaration of David Peters, Ph.D. in Support of

Petition for Inter Partes Review of U.S. Pat. No. 8,682,592", Jun. 13, 2019, 136 pages.

Illumina, Inc. v. Natera, Inc., "Order Re: Claim Construction", Jan. 30, 2019, 16 pgs.

Imielinski, M. et al., "Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing", Cell, vol. 150, Sep. 14, 2012,

Ingman, et al., "SNP Frequency Estimation Using Massively Parallel Sequencing of Pooled DNA", European Journal of Human

Genetics, vol. 17, No. 3, Oct. 15, 2008, 383-386. Innan, H. et al., "The Pattern of Polymorphism on Human Chromosome 21", Genome Research, vol. 13, 2003, 1158-1168. Interewicz, B. et al., "DNA Released from Ischemic and Rejecting

Organs as an Indicator of Graft Cellular Damage", Annals of Transplantation, vol. 9, No. 2, May 1, 2004, 42-45.

International Human, Genome Sequencing Consortium, "Finishing the Euchromatic Sequence of the Human Genome", Nature, vol. 431, Oct. 21, 2004, 931-945.
Ishii, et al., "Optimization of Annealing Temperature to Reduce

Bias Caused by a Primer Mismatch in Multitemplate PCR", Applied

and Environmental Microbiology, Aug. 2001, p. 3753-3755. Iskow, R. C. et al., "Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons", Cell, vol. 141, Jun. 25, 2010,

Ivanov, M. et al., "Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation", BMC Genomics, vol. 16 (Suppl 13):S1, Jun. 2015, 12 pgs.

Jabara, C. B. et al., "Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID", PNAS, vol. 108, No. 50, Dec. 13, 2011, 20166-20171.

Jahr, S. et al., "DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells", Cancer Research, vol. 61, Feb. 15, 2001,

Jamal-Hanjani, M. et al., "Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell Tung cancer", Annals of Oncology, vol. 27, No. 5, Jan. 28, 2016, 862-867.

Jamal-Hanjani, M. et al., "Tracking Genomic Cancer Evolution for Precision Medicine: The Lung TRACERx Study", PLOS Biology, vol. 12, No. 7, Jul. 2014, 1-7.

Jamal-Hanjani, M. et al., "Tracking the Evolution of Non-Small-Cell Lung Cancer", The New England Journal of Medicine, vol. 376, No. 22, Jun. 1, 2017, 2109-2121.

Jarvie, T., "Next generation sequencing technologies", Drug Discovery Today: Technologies, vol. 2, No. 3, 2005, 255-260. Jen, J. et al., "An Overview on the Isolation and Analysis of

Circulating Tumor DNA in Plasma and Serum", Annals New York Academy of Sciences, 2000, 8-12.

Jenkins, S. et al., "High-throughput SNP genotyping", Comparative and Functional Genomics, vol. 3, Dec. 5, 2001, 57-66.

Jennings, C. et al., "Investigation of Effects of Acid Citrate Dextrose and EDTA on Ability to Quantitatively Culture Human Immunodeficiency Virus", Journal of Clinical Microbiology, vol. 38, No. 9, 2000, 3522

Jett, K. et al., "Clinical and genetic aspects of neurofibromatosis 1", Genetics In Medicine, vol. 12, No. 1, Jan. 2010, 11 pages.

Jewesburty, E.C.O., "Reactions after Transfusion of Stored Blood", The British Medical Journal, vol. 1, No. 4191, 1941, 664-665. Jiang, P. et al., "The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics", Trends in Genetics,

Johnson, D. S. et al., "Comprehensive Analysis of Karyotypic Mosaicism."

Johnson, D. S. et al., "Genome-Wide Mapping of in Vivo Protein-DNA Interactions", Science, vol. 316, Jun. 8, 2007, 1497-1502.

Johnson, D.S. et al., "Comprehensive Analysis of Karyotypic Mosaicism."

Between Trophectoderm and Inner Cell Mass", Molecular Human Reproduction, 16(12), 2010, 944-949.

Johnson, J. B. et al., "Differential mechanisms of complementmediated neutralization of the closely related paramyxoviruses simian virus 5 and mumps virus", Virology, vol. 376, No. 1, 2008, 112-123. Johnson, K. L. et al., "Interlaboratory Comparison of Fetal Male DNA Detection from Common Maternal Plasma Samples by Real-Time PC", Clinical Chemistry, vol. 50, No. 3, 2004, 516-521

Johnson D.S, et al., "Preclinical Validation of a Microarray Method for Full Molecular Karyotyping of Blastomeres in a 24-h Protocol", Human Reproduction, 25 (4), 2010, 1066-1075.

Jordens, et al., "Amplification with molecular beacon primers and reverse line blotting for the detection and typing of human papillomaviruses", Journal of Virological Methods, vol. 89, 2000, 29-37. Jung, K. et al., "Cell-free DNA in the blood as a solid tu1nor biomarker—A critical appraisal of the literature", Clinica Chimica Acta, vol. 411, 2010, 1611-1624.

Juppner, H. et al., "Functional Properties of the PTH/PTHrP Receptor", Bone, vol. 17, No. 2 Supplement, Aug. 1995, 39S-42S.

Kaboev, et al., "PCR hot start using primers with the structure of molecular beacons (hairpin-like structure)", Nucleic Acids Research, vol. 28, 2000, 1-2.

Kalendar, Ruslan et al., "Java Web Tools for PCR, in Silico PCR, and Oligonucleotide Assembly and Analysis", Genomics, vol. 98, 2011, pp. 137-144.

Kamat, A. A. et al., "Quantification of total plasma cell-free DNA in ovarian cancer using real-time PCR", Ann N Y Acad Sci., vol.

1075, Sep. 2006, 230-234. Kamel, A. M. et al., "A simple strategy for breakpoint fragment determination in chronic myeloid leukemia", Cancer Genetics and Cytogenetics, vol. 122, 2000, 110-115.

Kane, M. et al., "Application of Less Primer Method to Commercial Kits", Forensic Science International: Genetics Supplement Series, vol. 1, Issue 1, 2008, 41-43.

Kane, M., "Application of Less Primer Method to Multiplex PCR", International Congress Series, vol. 1288, 2006, pp. 694-696.

Kanou, et al., "Cell-free DNA in human ex vivo lung perfusate as a potential biomarker to predict the risk of primary graft dysfunction in lung transplantation", The Journal of Heart and Lung Transplantation, vol. 36, No. 45, 2017, S187

Kapadia, Samir R. et al., "Impact of Intravascular Ultrasound in Understanding Transplant Coronary Artery Disease", Current Opinion In Cardiology, vol. 14, No. 2, Mar. 1, 1999, 1-19. Kaplinski, Lauris et al., "MultiPLX: Automatic Grouping and

Evaluation of PCR Primers", Bioinformatics, 21(8), 2005, 1701-

# US 11,530,454 B2

Page 17

#### (56) References Cited

#### OTHER PUBLICATIONS

Karger, et al., "DNA Sequencing by Capillary Electrophoresis", Electrophoresis, vol. 30, Supplement 1, Jun. 1, 2009, 1-11. Karoui, Noureddine E. et al., "Getting More from Digital SNP Data", Statistics in Medicine, vol. 25, Issue 18, Jan. 5, 2006, 3124, 313

Kass, et al., "Diagnosis of Graft Coronary Artery Disease", Current Opinion in Cardiology, vol. 22, No. 2, Mar. 1, 2007, 139-145. Kathiresan, Sekar et al., "Genome-wide Association of Early-onset Myocardial Infarction With Common Single Nucleotide Polymorphisms, Common Copy Number Variants, and Rare Copy Number Variants", Nature Genetics, vol. 41, No. 3, Mar. 1, 2009, 1-23.

Kazakov, V.I. et al., "Extracellular DNA in the Blood of Pregnant Women", Tsitologia, vol. 37, No. 3, 1995, 1-8.

Keith, L. et al., "Clinical Experience With the Prevention of Rh-Isoimmunization: A Historical Comparative Analysis", American Journal of Reproductive Immunology, vol. 5, 1984, 84-89. Keller, M. C. et al., "Non-Pathological Paternal Isodisomy of

Keller, M. C. et al., "Non-Pathological Paternal Isodisomy of Chromosome 2 Detected From a Genome-Wide SNP Scan", American Journal of Medical Genetics, Part A, 2009, 1823-1826.

Kennedy, S. R. et al., "Detecting ultralow-frequency mutations by Duplex Sequencing", Nature Protocols, vol. 9, No. 11, 2014, 2586-2606.

Keshavjee, S. H. et al., "The role of dextran 40 and potassium in extended hypothermic lung preservation for transplantation", The Journal of Thoracic and Cardiovascular Surgery, vol. 103, No. 2, 1992.

Kibbe, Warren A., "Oligocalc: An Online Oligonucleotide Properties Calculator", Nucleic Acids Research, vol. 35, 2007, pp. W43-W46

Kiernan, J. A., "Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: What they are and what they do.", Microscopy Today, vol. 1, 2000, 8-12.

Kijak, G. et al., "Discrepant Results in the Interpretation of HIV-1 Drug-Resistance Genotypic Data Among Widely Used Algorithms", HIV Medicine, 4, 2003, 72-78.

Kim, H. et al., "Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution", Genome Research, vol. 25, No. 3, Feb. 3, 2015, 316-327.

Kimmel, G. et al., "GERBIL: Genotype resolution and block identification using likelihood", PNAS, vol. 102, No. 1, Jan. 4, 2005, 158-162.

Kinde, I. et al., "Detection and quantification of rare mutations with massively parallel sequencing", PNAS, vol. 108, No. 23, Jun. 7, 2011, 9530-9535.

Kinnings, S. L. et al., "Factors affecting levels of circulating cell-free fetal DNA in maternal plasma and their implications for noninvasive prenatal testing", Prenatal Diagnosis, vol. 35, 2015, 816-822.

Kircher, Martin et al., "Improved Base Calling for the Illumina Genome Analyzer Using Machine Learning Strategies", Genome Biology, vol. 10, Issue 8, Article No. R83, Aug. 14, 2009, 83.2-83.9. Kirkizlar, E. et al., "Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology", Translational Oncology, vol. 8, No. 5, Oct. 2015, pp. 407-416.

Kirkness, E. F. et al., "Sequencing of isolated sperm cells for direct haplotyping of a human genome", Genome Research, vol. 23, 2013, 826-832.

Kittler, R. et al., "A Whole Genome Amplification Method to Generate Long Fragments from Low Quantities of Genomic DNA", Analytical Biochemistry, vol. 300, 2002, 237-244.

Kivioja, T. et al., "Counting absolute number of molecules using unique molecular identifiers", Nature Proceedings, Apr. 14, 2011, 18 pgs.

Kivioja, T. et al., "Counting absolute Nos. of molecules using unique molecular identifiers", Nature Methods, Advance Online Publication, Nov. 20, 2011, 1-5.

Kivioja, T. et al., "Counting absolute numbers of molecules using unique molecular identifiers", Nature Methods, vol. 9, No. 1, Jan. 2012, 72-76.

Kobashigawa, et al., "Multicenter Intravascular Ultrasound Validation Study Among Heart Transplant Recipients", Journal of the American College of Cardiology, vol. 45, No. 9, May 3, 2005, 1532-1537.

Koboldt, et al., "VarScan: Variant Detection in Massively Parallel Sequencing of Individual and Pooled Samples", Bioinformatics, vol. 25, No. 17, Jun. 19, 2009, 2283-2285.

Koelman, et al., "Donor-derived Soluble HLA Plasma Levels Can Not Be Used to Monitor Graft Rejection in Heart Transplant Recipients", Transplant Immunology, vol. 8, No. 1, Mar. 1, 2000, 57-64.

Kohler, C. et al., "Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors", Molecular Cancer, vol. 8, No. 105, Nov. 17, 2009, 9 pages.

Koide, K. et al., "Fragmentation of cell-free fetal DNA in plasma and urine of pregnant women", Prenatal Diagnosis, vol. 25, 2005, 604-607.

Koldehoff, Michael et al., "Quantitative analysis of chimerism after allogeneic stem cell transplantation by real-time polymerase chain reaction with single nucleotide polymorphisms, standard tandem repeats, and Y-chromosome-specific sequences", American Journal of Hematology, vol. 81, No. 10, Jul. 12, 2006, 735-746.

Konfortov, B A. et al., "A High-Resolution HAPPY Map of Dictyostelium discoideum Chromosome 6", Genome Research, vol. 10, No. 11, Nov. 2000, 1737-1742.

Konfortov, Bernard A. et al., "An Efficient Method for Multi-Locus Molecular Haplotyping", Nucleic Acids Research, 35(1), e6, 2007, 8 pgs.

Kopreski, Ms et al., "Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer", British Journal of Cancer, vol. 76, No. 10, 1997, 1293-1299.

Koressaar, Triinu et al., "Enhancements and Modifications of Primer Design Program Primer3", Bioinformatics, vol. 23, 2007, pp. 1289-1291

Korn, et al., "Integrated Genotype Calling and Association Analysis of SNPS, Common Copy Number Polymorphisms and Rare CNVS", Nature Genetics, vol. 40, No. 10, Oct. 1, 2008, 1253-1260.

Krjutskov, K. et al., "Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays", Nucleic Acids Research, vol. 36, No. 12, May 23, 2008, 7 pages.

Kuhn, H. et al., "Rolling-circle amplification under topological constraints", Nucleic Acids Research, vol. 30, No. 2, 2002, 574-580. Kukita, Y. et al., "High-fidelity target sequencing of individual molecules identified using barcode sequences: de nova detection and absolute quantitation of mutations in plasma cell-free DNA from cancer patients", DNA Research, vol. 22, No. 4, Jun. 29, 2015, 269-277.

Kuliev, Anver et al., "Thirteen Years' Experience on Preimplantation Diagnosis: Report of the Fifth International Symposium on Preimplantation Genetics", Reproductive BioMedicine Online, 8, 2, 2004, 229-235.

Kulifaj, D. et al., "Development of a standardized real time PCR for Torque teno viruses (TTV) viral load detection and quantification: A new tool for immune monitoring", Journal of Clinical Virology, vol. 105, 2018, 118-127.

Kumar, P. et al., "Ethylenegycol-Bis-(B-Aminoethylether)Tetraacetate as a Blood Anticoagulant: Preservation of Antigen-Presenting Cell Function and Antigen-Specific Proliferative Response of Peripheral Blood Mononuclear Cells from Stored Blood", Clinical and Diagnostic Laboratory Immunology, vol. 7, No. 4, 2000, 578-583.

Kunishima, S. et al., "First description of somatic mosaicism in MYH9 disorders", British Journal of Haematology, vol. 128, 2005, 360-365.

Kwok, P. Y., "High-throughput genotyping assay approaches", Pharmacogenomics, vol. 1, No. 1, 2000, 1-5. Lajoie, B. R. et al., "The Hitchhiker's Guide to Hi-C Analysis:

Lajoie, B. R. et al., "The Hitchhiker's Guide to Hi-C Analysis: Practical guidelines", Methods: Author manuscript, vol. 72, Jan. 2015, 65-75.

# US 11,530,454 B2

Page 18

### (56) References Cited

#### OTHER PUBLICATIONS

Lambert, et al., "Quantification of Maternal Microchimerism by HLA-Specific Real-time Polymerase Chain Reaction", Arthritis and Rheumatism, vol. 50, No. 3, Mar. 1, 2004, 906-914.

Lambert-Messerlian, G. et al., "Adjustment of Serum Markers in First Trimester Screening", Journal of Medical Screening, 16 (2), 2009, 102-103.

Landegren, U. et al., "Padlock and proximity probes for in situ and array-based analyses: tools for the post-genomic era", Comparative and Functional Genomics, vol. 4, 2003, 525-530.

Landegren, U. et al., "Reading Bits of Genetic Information: Methods for Single-Nucleotide Polymorphism Analysis", Genome Research, vol. 8, No. 8, 769-776, 1997.

Lander, E. S. et al., "Initial sequencing and analysis of the human genome", Nature, vol. 409, Feb. 15, 2001, 860-921.

Langmore, J., "Quality Control and Pre-Qualifications of NGS Libraries Made from Clinical Samples", ABRF 2013 Satellite Workshop, Mar. 2, 2013, 35 pages.

Lanman, et al., "Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA", Plos One, DOI:10.1371/journal.pone.0140712, 2015, 1-27.

Lapaire, O. et al., "Array-CGH analysis of cell-free fetal DNA in 10 mL of amniotic fluid supernatant", Prenatal Diagnosis, vol. 27, May 17, 2007, 616-621.

Lapierre, J.M. et al., "Analysis of uncultured amniocytes by comparative genomic hybridization: a prospective prenatal study", Prenatal Diagnosis, vol. 20, 2000, 123-131.

Lardeux, Frederic et al., "Optimization of a Semi-nested Multiplex PCR to Identify Plasmodium Parasites in Wild-Caught Anopheles in Bolivia, and Its Application to Field Epidemiological Studies", Transactions of the Royal Society of Tropical Medicine and Hygiene, vol. 102, 2008, pp. 485-492.

vol. 102, 2008, pp. 485-492. Larsen, J. B. et al., "Single-step Nested Multiplex PCR to Differentiate Between Various Bivalve Larvae", Marine Biology, vol. 146, 2005, pp. 1119-1129.

Lasken, R. S. et al., "Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens", TRENDS in Biotechnology, vol. 21, No. 12, Dec. 2003, 531-535. Lathi, Ruth B. et al., "Informatics Enhanced SNP Microarray Analysis of 30 Miscarriage Samples Compared to Routine Cytogenetics", PLoS ONE, 7(3), 2012, 5 pgs.

Lavebrat, et al., "Single Nucleotide Polymorphism (SNP) Allele

Lavebrat, et al., "Single Nucleotide Polymorphism (SNP) Allele Frequency Estimation in DNA Pools Using Pyrosequencing", Nature Protocols, vol. 1, No. 6, Jan. 11, 2007, 2573-2582

Protocols, vol. 1, No. 6, Jan. 11, 2007, 2573-2582. Lavebratt, Catharina et al., "Pyrosequencing-based SNP Allele Frequency Estimation in DNA Pools", Human Mutation, vol. 23, Issue 1, Dec. 19, 2003, 92-97.

Lavrentieva, I et al., "High Polymorphism Level of Genomic Sequences Flanking Insertion Sites of Human Endogenous Retroviral Long Terminal Repeats", FEBS Letters, vol. 443, No. 3, Jan. 29, 1999, 341-347.

Leamon, John H. et al., "A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions", Electrophoresis, vol. 24, No. 21, Nov. 1, 2003, 3769-3777. Leary, R. J. et al., "Development of Personalized Tumor Biomarkers

Leary, R. J. et al., "Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing", Science Translational Medicine, vol. 2, No. 20, Feb. 24, 2010, 1-8.

Leary, Rebecca J. et al., "Detection of Chromosomal Alterations in the Circulation of Cancer Patients with Whole-Genome Sequencing", Science Translational Medicine, 4, 162, 2012, 12.

Lecomte, T. et al., "Detection of Free-Circulating Tumor-Associated DNA in Plasma of Colorectal Cancer Patients and Its Association With Prognosis", Int. J. Cancer, vol. 100, 2002, 542-

Lee, et al., "ERBB2 kinase domain mutation in the lung squamous cell carcinoma", Cancer Letters, vol. 237, 2006, 89-94. Lee, J et al., "Anchored Multiplex PCR Enables Sensitive and

Lee, J et al., "Anchored Multiplex PCR Enables Sensitive and Specific Detection of Variants in Circulating Tumor DNA by Next-Generation Sequencing", DOI:https://doi.org/10.1016/j.cancergen. 2017.04.049, Cancer Genetics 214-215, 2017, 47.

Lee, T. et al., "Down syndrome and cell-free fetal DNA in archived maternal serum", AmJ Obstet Gynecol, vol. 187, No. 5, 1217-1221, Nov. 2002.

Lee, T.H. et al., "Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma", Transfusion, vol. 41, Feb. 2001, 276-282.

Levsky, J. M. et al., "Fluorescence in situ hybridization: past, present and future", Journal of Cell Science, vol. 116, No. 14, 2003, 2833-2838

Levsky, Jeffrey M. et al., "Efficacy of Coronary Ct Angiography Where We Are, Where We Are Going and Where We Want to Be", Journal of Cardiovascular Computed Tomography, vol. 3, Supplement 2, Nov. 2, 2009, s99-s108.

Li, et al., "Detection of SNPs in the Plasma of Pregnant Women and in the Urine of Kidney Transplant Recipients by Mass Spectrometry", Annals of the New York Academy of Sciences, vol. 1075, Sep. 5, 2006, 144-147.

Li, et al., "Mapping Short DNA Sequencing Reads and Calling Variants Using Mapping Quality Scores", Genome Research, vol. 18, No. 11,, Aug. 19, 2008, 1851-1858.

Li, et al., "Multiplex Padlock Targeted Sequencing Reveals Human Hypermutable CpG Variations", Genome Research, vol. 19, No. 9, Jun. 12, 2009, 1606-1615.

Li, et al., "SOAP2: An Improved Ultrafast Tool for Short Read Alignment", Bioinformatics, vol. 25, No. 15, Aug. 1, 2009, 1966-1967.

Li, B., "Highly Multiplexed Amplicon Preparation for Targeted Re-Sequencing of Sample Limited Specimens Using the Ion AmpliSeq Technology and Semiconductor Sequencing", Proceedings of the Annual Meeting of the American Society of Human Genetics [retrieved on Oct. 30, 2012], Retrieved from the Internet: <URL http://www.ashg.org/2012meeting/abstracts/fulltext/f120121811. htm>, 2012, 1 pg.

Li, R. et al., "SNP detection for massively parallel whole-genome resequencing", Genome Research, vol. 19, 2009, 1124-1132.

Li, Y. et al., Detection of Paternally Inherited Fetal Point Mutations for b-Thalassemia Using Size-Fractionated Cell-Free DNA in Maternal Plasma", JAMA, vol. 293, No. 7, Apr. 13, 2005, 843-849.

nal Plasma", JAMA, vol. 293, No. 7, Apr. 13, 2005, 843-849. Li, Y. et al., "Non-Invasive Prenatal Diagnosis Using Cell-Free Fetal DNA in Maternal Plasma from PGD Pregnancies", Reproductive BioMedicine Online, 19 (5), 2009, 714-720.

Li, Ying et al., "Detection of Donor-specific DNA Polymorphisms in the Urine of Renal Transplant Recipients", Clinical Chemistry, vol. 49, No. 4, Apr. 1, 2003, 655-658.

vol. 49, No. 4, Apr. 1, 2003, 655-658. Li, Ying et al., "Ready detection of donor-specific single-nucleotide polymorphisms in the urine of renal transplant recipients by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", Clin Chern, Oct. 2005, vol. 51, lssue.10,pp. 1903-1904, Oct. 1, 2005, 1903-1904.

Li, Ying et al., "Size Separation of Circulatory DNA in Maternal Plasma Permits Ready Detection of Fetal DNA Polymorphisms", Clinical Chemistry, 50, 6, 2004, 1002-1011.
Liao, Gary J.W. et al., "Targeted Massively Parallel Sequencing of

Liao, Gary J.W. et al., "Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles", Clinical Chemistry, 57 (1), 2011,92-101.

Liao, J. et al., "An Alternative Linker-Mediated Polymerase Chain Reaction Method Using a Dideoxynucleotide to Reduce Amplification Background", Analytical Biochemistry 253, 137-139 (1997). Lichtenstein, A. V. et al., "Circulating Nucleic Acids and Apoptosis", Annals New York Academy of Sciences, vol. 945, Aug. 1, 2001, 239-249.

Liew, Michael et al., "Genotyping of Single-Nucleotide Polymorphisms", Clinical Chemistry, 50(7), 2004, 1156-1164. Life Technologies, "Ion AmpliSeq Comprehensive Cancer Panel", 2012, 2 pgs.

Life Technologies, "Ion AmpliSeq<sup>TM</sup> Designer provides full flexibility to sequence genes of your choice", 2012, 4 pages.

Liljedahl, Ulrika et al., "Detecting Imbalanced Expression of SNP Alleles by Minisequencing on Microarrays", BMC Biotechnology, vol. 4, Article No. 24, Oct. 22, 2004, 1-10.

# US 11,530,454 B2

Page 19

### (56) References Cited

#### OTHER PUBLICATIONS

Lindberg, J. et al., "Exome Sequencing of Prostate Cancer Supports the Hypothesis of Independent Tumour Origins", European Urology, vol. 63, 2013, 347-353.

Lindroos, Katatina et al., "Genotyping SNPs by Minisequencing Primer Extension Using Oligonucleotide Microarrays", Methods in Molecular Biology, 212, Single Nucleotide Polymorphisms: Methods and Protocols, P-K Kwok (ed.), Humana Press, Inc., Totowa, NJ, 2003, 149-165.

Lo, et al., "Digital PCR for the Molecular Detection of Fetal Chromosomal Aneuploidy", PNAS, vol. 104, No. 32, Aug. 7, 2007, 13116-13121.

Lo, et al., "Fetal Nucleic Acids in Maternal Blood: the Promises", Clin. Chem. Lab. Med., 50(6), 2012, 995-998.

Lo, et al., "Free Fetal DNA in Maternal Circulation", JAMA, 292(23), (Letters to the Editor), 2004, 2835-2836.

Lo, et al., "Next-generation Sequencing of Plasma/Serum DNA: An Emerging Research And Molecular Diagnostic Tool", Clinical Chemistry, vol. 55, No. 4, Apr. 1, 2009, 607-608.

Lo, "Non-Invasive Prenatal Diagnosis by Massively parallel Sequencing of Maternal Plasma DNA", Open Biol 2: 120086, 2012, 1-5. Lo, et al., "Prenatal Sex Determination by DNA Amplification from Maternal Peripheral Blood", The Lancet, 2, 8676, 1989, 1363-1365. Lo, et al., "Presence of Donor-specific DNA in Plasma of Kidney and Liver-transplant Recipients", Lancet, vol. 351, No. 9112, May 2, 1998, 1329-1330.

Lo, et al., "Rapid Clearance of Fetal DNA from Maternal Plasma", Am. J. Hum. Genet., 64, 1999, 218-224.

Lo, et al., "Strategies for the Detection of Autosomal Fetal DNA Sequence from Maternal Peripheral Blood", Annals New York Academy of Sciences,731, 1994, 204-213.

Lo, et al., "Two-way cell traffic between mother and fetus: biologic

Lo, et al., "Two-way cell traffic between mother and fetus: biologic and clinical implications", Blood, 88(11), Dec. 1, 1996, 4390-4395. Lo, Y M. et al., "Circulating Nucleic Acids in Plasma and Serum: An Overview", Annals of the New York Academy of Sciences, vol. 945, Sep. 1, 2001, 1-7.

Lo, Y., "Noninvasive prenatal detection of fetal chromosomal aneuploidies by maternal plasma nucleic acid analysis: a review of the current state of the art", BJOG An International Journal of Obstetrics and Gynaecology, vol. 116, 2009, 152-157.

Lo, Y.M. D. et al., "Prenatal Diagnosis of Fetal RhD Status by Molecular Analysis of Maternal Plasma", The New England Journal of Medicine, vol. 339, No. 24, 1998, 1734-1738.

Lo, Y.M. Dennis, "Fetal Nucleic Acids in Maternal Plasma: Toward the Development of Noninvasive Prenatal Diagnosis of Fetal Chromosomal Aneuploidies", Ann. N.Y. Acad. Sci., 1137, 2008, 140-143.

Lo, Y.M. Dennis et al., "Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus", Science Translational Medicine, 2 (61), 2010, 13.

Lo, Y.M. Dennis et al., "Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection", Nature Medicine, 13 (2), 2007, 218-223.

Lo, Y.M. Dennis et al., "Presence of Fetal DNA in Maternal Plasma and Serum", The Lancet, 350, 1997, 485-487.

Lo, Y.M. Dennis et al., "Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis", Am. J. Hum. Genet., 62, 1998, 768-775.

Lo, Y.M.D., "Fetal DNA in Maternal Plasma: Biology and Diagnostic Applications", Clinical Chemistry, vol. 46, No. 12, 2000, 1903-1906.

Lo, Y.M.D. et al., "Prenatal diagnosis: progress through plasma nucleic acids", Nature Reviews, vol. 8, 2007, 71-77.

Lo, Y-M D., "Non-invasive prenatal diagnosis using fetal cells in

Lo, Y-M D., "Non-invasive prenatal diagnosis using fetal cells in maternal blood", J. Clin. Pathol., vol. 47, 1994, 1060-1065.

Lo, Y-M.D et al., "Detection of Single-Copy Fetal DNA Sequence from Maternal Blood", The Lancet, 335, 1990, 1463-1464.

Lo, Y-M.D et al., "Prenatal Determination of Fetal Rhesus D Status by DNA Amplification of Peripheral Blood of Rhesus-Negative Mothers", Annals New York Academy of Sciences, 731, 1994, 229-236.

Lo, Y-M.D. et al., "Detection of Fetal RhD Sequence from Peripheral Blood of Sensitized RhD-Negative Pregnant Women", British Journal of Haematology, 87, 1994, 658-660.

Lo, Y-M.D. et al., "Prenatal Determination of Fetal RhD Status by Analysis of Peripheral Blood of Rhesus Negative Mothers", The Lancet, 341, 1993, 1147-1148.

Loh, Elwyn, "Anchored PCR: Amplification with Single-sided Specificity", Methods, vol. 2, 1991, pp. 11-19.

Lovmar, L. et al., "Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DN", Nucleic Acids Research, vol. 31, No. 21, 2003,, 9 pgs.

Lu, I. et al., "Establishment of a system based on universal multiplex-PCR for screening genetically modified crops", Anal. Bioanal. Chem., vol. 396, Oct. 24, 2009, 2055-2064.

Lu, S. et al., "Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing", Science, vol. 338, Dec. 21, 2012, 1627-1630.

Lui, Y. Y. et al., "Predominant Hematopoietic Origin of Cell-Free DNA in Plasma and Serum after Sex-Mismatched Bone Marrow Transplantation", Clinical Chemistry, vol. 48, vol. 3, 2002, 421-427. Lui, Yanni Y. et al., "Circulating DNA in Plasma and Serum: Biology, Preanalytical Issues and Diagnostic Applications", Clinical Chemistry and Laboratory Medicine, vol. 40, No. 10, Oct. 29, 2002, 962-968

Lui, Yanni Y. et al., "Origin of Plasma Cell-Free DNA after Solid Organ Transplantation", Clinical Chemistry, vol. 49, No. 3, Mar. 1, 2003. 495-496.

Lun, Fiona M. et al., "Microfluidics Digital PCR Reveals A Higher Than Expected Fraction of Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 54, No. 10, Aug. 14, 2008, 1664-1672. Lun, Fiona M. et al., "Noninvasive Prenatal Diagnosis of Monogenic

Lun, Fiona M. et al., "Noninvasive Prenatal Diagnosis of Monogenic Diseases by Digital Size Selection and Relative Mutation Dosage on DNA in Maternal Plasma", PNAS, 105(50), 2008, 19920-19925.

Ma, Xiaotu et al., "Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia", Nature Communications, vol. 6, Mar. 19, 2015, 1-12.

Mackiewicz, D. et al., "Distribution of Recombination Hotspots in the Human Genome—A Comparison of Computer Simulations with Real Data" PLOS One, vol. 8, No. 6, Jun. 2013, 11 pages

Real Data", PLOS One, vol. 8, No. 6, Jun. 2013, 11 pages. Magbanua, M. J. et al., "Abstract PD2-01: Personalized serial circulating tumor DNA (ctDNA) analysis in high-risk early stage breast cancer patients to monitor and predict response to neoadjuvant therapy and outcome in the I-Spy 2 Trial", Cancer Research, vol. 79, No. 4 Supplement, Feb. 15, 2019.

Mamon, H. et al., "Letters to the Editor: Preferential Amplification

Mamon, H. et al., "Letters to the Editor: Preferential Amplification of Apoptotic DNA from Plasma: Potential for Enhancing Detection of Minor DNA Alterations in Circulating DNA", Clinical Chemistry, vol. 54, No. 9, 2008, 1582-1584.

Maniatis, T. et al., "In: Molecular Cloning: A Laboratory Manual",

Maniatis, T. et al., "In: Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory, New York, Thirteenth Printing, 1986, 458-459.

Mansfield, Elaine S, "Diagnosis of Down Syndrome and Other Aneuploidies Using Quantitative Polymerase Chain Reaction and Small Tandem Repeat Polymorphisms", Human Molecular Genetics, 2, 1, 1993, 43-50.

Mardis, E. R., "The impact of next-generation sequencing technol-

Mardis, E. R., "The impact of next-generation sequencing technology on genetics", Trends in Genetics, vol. 24, No. 3, Feb. 11, 2008, 133-141.

Marguiles, M. et al., "Genome Sequencing in Open Microfabricated High Density Picoliter Reactors", Nature, vol. 437, No. 7057, Sep. 15, 2005, 376-380.

Margulies, M. et al., "Genome sequencing in microfabricated high-density picolitre reactors", Nature, vol. 437, Sep. 15, 2005, 376-380. Margulies, M. et al., "Genome sequencing in microfabricated high-density picolitre reactors plus Supplemental Methods", Nature, vol. 437, Sep. 15, 2005, 40 pgs.

# US 11,530,454 B2

Page 20

#### (56) References Cited

#### OTHER PUBLICATIONS

Marianes, Alexis E. et al., "Targets of Somatic Hypermutation within Immunoglobulin Light Chain Genes in Zebrafish", Immunology, vol. 132, 2010, pp. 240-255.

Markoulatos, P. et al., "Multiplex Polymerase Chain Reaction: A Practical Approach", Journal of Clinical Laboratory Analysis, vol. 16, 2002, 47-51.

Maron, Jill L. et al., "Cell-free Fetal DNA Plasma Extraction and Realtime Polymerase Chain Reaction Quantification", Methods in Molecular Medicine, vol. 132, Aug. 1, 2007, 51-63.

Marshutina, N. V. et al., "Comparative Clinical and Diagnostic Significance of Some Serological Tumor Associated Markers for Different Histological Types of Lung Cancer", Russian Oncological Journal, vol. 3, 2010, 13-16.

Martinez- Lopez, J. et al., "Real-time PCR Quantification of Haematopoietic Chimerism after Transplantation: A Comparison Between TaqMan And Hybridization Probes Technologies", International Journal of Laboratory Hematology, vol. 32, Issue 1, Part 1, May 12, 2009, e17-e25.

Martins, et al., "Quantification of Donor-derived DNA In Serum: A

Martins, et al., "Quantification of Donor-derived DNA In Serum: A New Approach of Acute Rejection Diagnosis in a Rat Kidney Transplantation Model", Transplantation Proceedings, vol. 37, No. 1,, Jan. 1, 2005, 87-88.

Masuzaki, H. et al., "Detection of cell free placental DNA in maternal plasma: direct evidence from three cases of confined placental mosaicism", J Med Genet, vol. 41,2004, 289-292. Matsubara, T. et al., "Pantropic Retroviral Vectors Integrate and

Matsubara, T. et al., "Pantropic Retroviral Vectors Integrate and Express in Cells of the Malaria Mosquito, Anopheles Gambiae", PNAS, vol. 93, 1996, pp. 6181-6185.

Matsuzaki, H et al., "Genotyping over 100,000 SNPs on a pair of

Matsuzaki, H et al., "Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays", Nature Methods, vol. 1, No. 2, Nov. 2004, 109-111.

May, Robert M., "How Many Species Are There on Earth?", Science, 241, Sep. 1988, 1441-1449.

McBride, D. et al., "Use of Cancer-Specific Genomic Rearrangements to Quantify Disease Burden in Plasma from Patients with Solid Tumors", Genes, Chromosomes & Cancer, vol. 49, Aug. 19, 2010, 1062-1069.

McCloskey, M. L. et al., "Encoding PCR Products with Batch-stamps and Barcodes", Biochem Genet., vol. 45, Oct. 23, 2007, 761, 767

McCray, Alexa T et al., "Aggregating UMLS Semantic Types for Reducing Conceptual Complexity", MEDINFO 2001: Proceedings of the 10th World Congress on Medical Informatics (Studies in Health Technology and Informatics, 84, V. Patel et al. (eds.), IOS Press Amsterdam, 2001, 216-220.

McDonald, B. R. et al., "Abstract P4-01-21: Multiplexed targeted

McDonald, B. R. et al., "Abstract P4-01-21: Multiplexed targeted digital sequencing of circulating tumor DNA to detect minimal residual disease in early and locally advanced breast cancer", Cancer Research, vol. 79, No. 4 Supplement, Feb. 15, 2019. McDonald, J. P. et al., "Novel thermostable Y-family polymerases

McDonald, J. P. et al., "Novel thermostable Y-family polymerases applications for the PCR amplification of damaged or ancient DNAs", Nucleic Acids Research, vol. 34, No. 4, 2006, 1102-1111. Mennuti, M. et al., "Is It Time to Sound an Alarm About False-Positive Cell-Free DNA Testing for Fetal Aneuploidy?", American Journal of Obstetrics, 2013, 5 pgs.

Merriam-Webster, "Medical Definition of Stimulant", http://www.

Merriam-Webster, "Medical Definition of Stimulant", http://www.merriam-webster.com/medical/stimulant, Mar. 14, 2016, 7 pages. Merriam-Webster, "Universal Definition", Merriam-Webster.com, 2014, 3 pages.

Mersy, et al., "Noninvasive Detection of Fetal Trisomy 21: Systematic Review and Report of Quality and Outcomes of Diagnostic Accuracy Studies Performed Between 1997 and 2012", Human Reproduction Update, 19(4), 2013, 318-329.

Mertes, F. et al., "Targeted enrichment of genomic DNA regions for next-generation sequencing", Briefings in Functional Genomics, vol. 10, No. 6, Nov. 26, 2011, 374-386.

Messmer, Trudy O. et al., "Application of a Nested, Multiplex PCR to Psittacosis Outbreaks", Journal of Clinical Microbiology, vol. 35, No. 8, 1997, pp. 2043-2046.

Metzker, M. L. et al., "Polymerase Chain Reaction", Encyclopedia of Medical Devices and Instrumentation, vol. 5, Second Edition, 2006, 380-387.

Metzker, M. L. et al., "Quantitation of Mixed-Base Populations of HIV-1 Variants by Automated DNA Sequencing with BODIPY\* Dye-Labeled Primers", BioTechniques, vol. 25, Sep. 1998, 446-462.

Meuzelaar, Linda S. et al., "Megaplex PCR: A Strategy for Multiplex Amplification", Nature Methods, vol. 4, 2007, pp. 835-837. Meyer, M et al., "Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing", Cold Spring Harbor Protocols, vol. 2010, Issue 6, Jun. 2010, 1-10.

Meyerson, M. et al., "Advances in understanding cancer genomes through second-generation sequencing", Nature Reviews: Genetics, vol. 11, Oct. 2010, 685-696.

Mikkelsen, T. S. et al., "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells", Nature, vol. 448, No. 2, Aug. 2007, 553-562.

Milani, et al., "Genotyping Single Nucleotide Polymorphisms by Multiplex Minisequencing Using Tag-arrays", DNA Microarrays for Biomedical Research, vol. 529, Jan. 16, 2009, 215-229.

Miller, Robert, "Hyperglycemia-Induced Changes in Hepatic Membrane Fatty Acid Composition Correlate with Increased Caspase-3 Activities and Reduced Chick Embryo Viability", Comparative Biochemistry and Physiology, Part B, 141, 2005, 323-330.

Miller, Robert R., "Homocysteine-Induced Changes in Brain Membrane Composition Correlate with Increased Brain Caspase-3 Activities and Reduced Chick Embryo Viability", Comparative Biochemistry and Physiology Part B, 136, 2003, 521-532.

Miner, B. E. et al., "Molecular barcodes detect redundancy and

Miner, B. E. et al., "Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR", Nucleic Acids Research, vol. 32, No. 17, Sep. 30, 2004, 1-4.

Minkoff, E. et al., "Stem Cells, Cell Division, and Cancer", Biology

Minkoff, E. et al., "Stem Cells, Cell Division, and Cancer", Biolog Today Third Edition, Chapter 12, 2004, 10 pages.

Miramontes, Pedro et al., "DNA Dimer Correlations Reflect in Vivo Conditions and Discriminate Among Nearest-neighbor Base Pair Free Energy Parameter Measures", Physica A, vol. 321, 2003, pp. 577-586.

Mitra, S. et al., "Chapter 4 Classification Techniques", Introduction to Machine Learning and Bioinformatics, First Edition, 2008, 101-127

Morand, et al., "Hesperidin contributes to the vascular protective effects of orange juice: a randomized crossover study in healthy volunteers", Am J Clin Nutr. Jan. 2011;93(1):73-80. Epub Nov. 10, 2010.

Moreau, Valerie et al., "Zip Nucleic Acids: New High Affinity Oligonucleotides as Potent Primers for PCR and Reverse Transcription", Nucleic Acids Research, vol. 37, No. 19, e130, 2009, 14 pages.

Moreira, et al., "Increase in and Clearance of Cell-free Plasma DNA in Hemodialysis Quantified by Real-time PCR", Clinical Chemistry and Laboratory Medicine, vol. 44, No. 12, Dec. 13, 2006, 1410-1415

Morris, J. K. et al., "Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register", BMJ Online, vol. 339, Oct. 2009, 5 pages.

Munne, S. et al., "Chromosome Abnormalities in Human Embryos", Textbook of Assisted Reproductive Techniques, 2004, pp. 355-377. Munne, S. et al., "Chromosome abnormalities in human embryos", European Society of Human Reproduction and Embryology: Human Reproduction Update, vol. 4, No. 6, 1998, 842-855.

Munne, S. et al., "Improved implantation after preimplantation genetic diagnosis of aneuploidy", Reproductive BioMedicine Online, vol. 7., No. L. May 15, 2003, 91-97.

vol. 7., No. 1., May 15, 2003, 91-97.
Murali, R. et al., "Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: The Fab is directed against an intermediate in the helix-coil dynamics of the enzyme", Proc. Natl. Acad. Sci. USA, vol. 95, Oct. 1998, 12562-12567.

Murtaza, M. et al., "Non-Invasive Analysis of Acquired Resistance to Cancer Therapy by Sequencing of Plasma DNA", Nature (doi:10. 1038/nature12065), 2013, 6 pgs.

# US 11,530,454 B2

Page 21

#### (56) References Cited

#### OTHER PUBLICATIONS

Muse, Spencer V., "Examining rates and patterns of nucleotide substitution in plants", Plant Molecular Biology 42: 25-43, 2000. Myers, Chad L. et al., "Accurate Detection of Aneuploidies in Array CGH and Gene Expression Microarray Data", Bioinformatics, 20(18), 2004, 3533-3543.

Nagalla, S. R. et al., "Proteomic Analysis of Maternal Serum in Down Syndrome: Identification of Novel Protein Biomarkers", Journal of Proteome Research, vol. 6, Mar. 21, 2007, 1245-1257. Nakamura, N. et al., "Ex Vivo Liver Perfusion with Arterial Blood from a Pig with Ischemic Liver Failure", Artificial Organs, vol. 23, No. 2, 1999, 153-160.

Namlos, H. M. et al., "Noninvasive Detection of ctDNA Reveals Intratumor Heterogeneity and is Associated with Tumor Burden in Gastrointestinal Stromal Tumor", Molecular Cancer Therapeutics, vol. 17, No. 11, 2018, 2473-2480.

Namlos, H.M. et al., "Use of liquid biopsies to monitor disease progression in a sarcoma patient: a case report", BMC Cancer, vol. 17, No. 1, 2017, 2-3.

Nannya, Yasuhito et al., "A Robust Algorithm for Copy Number Detection Using High-density Oligonucleotide Single Nucleotide Polymorphism Genotyping Arrays", Cancer Res., 65, 14, 2005, 6071-6079.

Narayan, A. et al., "Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing", Cancer Research, vol. 72, No. 14, Jul. 15, 2012, 3492-3498.

Natera, Inc., "Declaration of Sandra L. Haberny", May 16, 2019, 3 pages.

Natera, Inc., "Defendant Natera, Inc.'s Invalidity Contentions Under Patent L.R. 3-3; Document Production Accompanying Invalidity Contentions Under Patent L.R. 3-4", Aug. 20, 2018, 17 pages.

Natera, Inc., "Exhibit 8 EHRICH Invalidity Chart", Aug. 20, 2018, 16 pages.

Natera, Inc., "Exhibits A-H to Haberny Declaration", May 16, 2019, 192 pages.

Natera, Inc., "Motion to Dismiss", May 16, 2019, 2 pages.

Natera, Inc., "Natera Inc.'s First Amended Answer, Affirmative Defenses and Counterclaims", Aug. 16, 2018, 28 pages.

Natera, Inc., "Natera, Inc.'s Supplemental Objections and Response to Plaintiff Illumina, Inc.'s Interrogatory No. 8", Mar. 20, 2019, 29 pages.

Natera, Inc., "Opening Brief in Support of Motion to Dismiss", May 16, 2019, 26 pages.

16, 2019, 26 pages.
Natera, Inc., "Petitioner Reply Per Board Order of Nov. 2, 2018 (Paper No. 10)", Nov. 9, 2018, 8 pgs.
National Institutes of Health, "Genetics Home Reference: Your

National Institutes of Health, "Genetics Home Reference: Your Guide to Understanding Genetic Conditions", Feb. 28, 2014, 2 pgs. Nawroz, H. et al., "Microsatellite Alterations in Serum DNA of Head and Neck Cancer Patients", Nature Medicine, vol. 2, No. 9, Sep. 1996, 1035-1037.

NCBI, "Blast of AAAAAAAAATTTAAAAAAAAATTT", http://blast.ncbi.nlm.nih.gov/Blast.cgi, 2015, 9 pages.

NCBI, "db SNP rs2056688", http://www.ncbi.nlm.nih.gov/proiects/SNP/snp\_ref.cgi?rs=2056688, 2015, 3 pages.

NCBI, "dbSNP record for rs1294331", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs 1294331 >, 2019, 2 pgs.

pgs. NCBI, "dbSNP record for rs1872575", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs1872575, 2019, 2 pgs. NCBI, "dbSNP record for rs2362450", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs2362450>, 2019, 1 pg. NCBI, "dbSNP record for rs2384571", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs2384571>, 2019, 2 pgs. NCBI, "dbSNP record for rs2498982", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs2498982>, 2019, 3 pgs. NCBI, "dbSNP record for rs3731877", Retrieved from the Internet: <URL: www.ncbi.nlm.nih.gov/snp/?term=rs3731877>, 2019, 2 pgs.

Nelson, C. M. et al., "Whole genome transcription profiling of Anaplasma phagocytohilum in human and tick host cells by tiling array analysis", BMC Genomics, vol. 9, No. 364, Jul. 31, 2008, 16 pgs.

Neve, B. et al., "Rapid SNP Allele Frequency Determination in Genomic DNA Pools by Pyrosequencing", BioTechniques, vol. 32, No. 5, May 1, 2002, 1138-1142.

New England Biolabs, "Nucleic Acids, Linkers and Primers: Random Primers", 1998/99Catalog, 1998, 121 and 284.

Newman, A. M. et al., "Integrated digital error suppression for improved detection of circulating tumor DNA", Nature Biotechnology, vol. 34, No. 5, May 2016, 547-555.

ogy, vol. 34, No. 5, May 2016, 547-555. Ng, et al., "Multiplex Sequencing of Paired-end Ditags (MS-PET): A Strategy for the Ultra-high-throughput Analysis of Transcriptomes and Genomes", Nucleic Acids Research, vol. 34, No. 12, Jul. 13, 2006, 1-10.

Ng, S. B. et al., "Individualised multiplexed circulating tumour DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery", Scientific Reports, vol. 7, No. 40737, Jan. 19, 2017, 11 pages.

Jan. 19, 2017, 11 pages. Nguyen-Dumont, T., "A high-plex PCR approach for massively parallel sequencing", BioTechniques, vol. 55, No. 2, Aug. 2013, 69-74

Nicolaides, K. et al., "Noninvasive Prenatal Testing for Fetal Trisomies in a Routinely Screened First-Trimester Population", American Journal of Obstetrics (article in press), 207, 2012, 1, e1-1 e6

Nicolaides, K.H. et al., "Validation of Targeted Sequencing of Single-Nucleotide Polymorphisms for Non-Invasive Prenatal Detection of Aneuploidy of Chromosomes 13, 18, 21, X, and Y", Prenatal Diagnosis, 33, 2013, 575-579.

Nicolaides, Kypros H. et al., "Prenatal Detection of Fetal Triploidy from Cell-Free DNA Testing in Maternal Blood", Fetal Diagnosis and Therapy, 2013, 1-6.

Nilsson, et al., "Analyzing genes using closing and replicating circles", Trends in Biotechnology, 24, 2006, 83-88.

Nilsson, M. et al., "Padlock Probes: Circularizing Oligonucleotides for Localized DNA Detection", Science, vol. 265, Sep. 10, 1994, 2085-2088.

Nishigaki, K. et al., "Random PCR-Based Genome Sequencing: A Non-Divide-and-Conquer Strategy", DNA Research, vol. 7, 2000, 19.26

Nishiwaki, Morie et al., "Genotyping of Human Papillomaviruses by a Novel One-step Typing Method With Multiplex PCR and Clinical Applications", Journal of Clinical Microbiology, vol. 46, 2008, pp. 1161-1168.

Norton, S. E. et al., "A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR", Clin Biochem., vol. 46, No. 15, Oct. 2013, 1561-1565.

Nui, A. et al., "The Functional Integrity of a Normothermic Perfusion System Using Artificial Blood in Pig Liver", Journal of Surgical Research, Vo. 131, 2006, 189-198.

Nygren, et al., "Quantification of Fetal DNA by Use of Methylation-Based DNA Discrimination", Clinical Chemistry 56:10 1627-1635 (2010).

O'connell, G. C. et al., "High Interspecimen Variability in Nucleic Acid Extraction Efficiency Necessitates the Use of Spike-In Control for Accurate qPCR-based Measurement of Plasma Cell-Free DNA Levels", Lab Medicine, vol. 48, 2017, 332-338.

Oeth, et al., "iPLEXTM Assay: Increased Plexing Efficiency and Flexibility for MassARRAY System Through Single Base Primer Extension with Mass-Modified Terminators", Sequenom Application Note Doc. No. 8876-006, Apr. 28, 2005, 1-12.

Ogino, S. et al., "Bayesian Analysis and Risk Assessment in Genetic Counseling and Testing", Journal of Molecular Diagnostics, 6 (1), 2004, 9 pgs.

Ohara, O et al., "One-sided Polymerase Chain Reaction: The Amplification of cDNA", Proceedings of the National Academy of Sciences, vol. 86, 1989, 5673-5677.

Ohira, T. et al., "Tumor volume determines the feasibility of cell-free DNA sequencing for mutation detection in non-small cell lung cancer", Cancer Science, vol. 107, 2016, 1660-1666.

# US 11,530,454 B2

Page 22

#### (56) References Cited

#### OTHER PUBLICATIONS

Ohsawa, M. et al., "Prenatal Diagnosis of Two Pedigrees of Fukuyama Type Congenital Muscular Dystrophy by Polymorphism Analysis", The Health and Welfare Ministry, 1994, 5 pgs.

Ohya, K. et al., "Detection of the CTG Repeat Expansion in Congenital Myotonic Dystrophy", Jpn J. Human Genet, vol. 42, 1997, 169-180.

Okou, et al., "Microarray-based Genomic Selection for High-throughput Resequencing", Nature Methods, vol. 4, No. 11, Oct. 14, 2007, 907-909.

Okou, David T. et al., "Combining Microarray-based Genomic Selection (MGS) with the Illumina Genome Analyzer Platform to Sequence Diploid Target Regions", Annals of Human Genetics, vol. 73, No. 5, Aug. 6, 2009, 502-513.

Olerup, O. et al., "HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation", Tissue Antigens, vol. 39, No. 5, May 1992, 225-235.

Oliphant, A. et al., "Bead.Array<sup>TM</sup> Technology: Enabling an Accu-

Oliphant, A. et al., "Bead.Array™ Technology: Enabling an Accurate, Cost-Effective Approach to High-Throughput Genotyping", Bio Techniques, vol. 32, Jun. 2002, S56-S6.

Olivarius, S. et al., "High-throughput Verification of Transcriptional starting Sites by Deep-RACE", Bio Techniques, vol. 46, No. 2, Feb. 2009, 130-132.

Olive, M. et al., "Characterization of the DiFi Rectal Carcinoma Cell Line Derived from a Familial Adenomatous Polyposis Patient", In Vitro Cellular & Developmental Biology, vol. 29A, No. 3, Part 1, Mar. 1993, 239-248.

Oliver, Dwight H. et al., "Use of Single Nucleotide Polymorphisms (SNP) and Real-time Polymerase Chain Reaction for Bone Marrow Engraftment Analysis", The Journal of Molecular Diagnostics, vol. 2, No. 4, Nov. 1, 2000, 202-208.

Olivier, et al., "The Invader Assay for SNP Genotyping", Mutation Research, vol. 573, No. 1-2, Jun. 3, 2005, 103-110. Olney, R. S. et al., "Chorionic Villus Sampling and Amniocentesis:

Olney, R. S. et al., "Chorionic Villus Sampling and Amniocentesis: Recommendations for Prenatal Counseling", MMWR: Recommendations and Reports, 44(RR-9), Jul. 21, 1995, 1-12.

O'Malley, R. et al., "An adapter ligation-mediated PCR method for

O'Malley, R. et al., "An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome", Nat. Protoc., 2, 2007, 2910-2917.

Orozco A.F., et al., "Placental Release of Distinct DNA-Associated Micro-Particles into Maternal Circulation: Reflective of Gestation Time and Precelampsia", Placenta, 30, 2009, 891-897.

Orsouw, et al., "Complexity Reduction of Polymorphic Sequences (Crops): A Novel Approach for Large-scale Polymorphism Discovery in Complex Genomes", PLoS ONE, vol. 11:e1172, Nov. 14, 2017, 1-10.

Owczarzy, Richard et al., "Melting Temperatures of Nucleic Acids: Discrepancies in Analysis", Biophysical Chemistry, vol. 117, 2005, pp. 207-215.

Ozawa, Makiko et al., "Two Families with Fukuyama Congenital Muscular Dystrophy that Underwent in Utero Diagnosis Based on Polymorphism Analysis", Clinical Muscular Dystrophy: Research in Immunology and Genetic Counseling—FY 1994 Research Report, (including copy of text in Japanese), 1994, 8.

Paez, Guillermo J. et al., "Genome coverage and sequence fidelity of  $\Phi$ 29 polymerase-based multiple strand displacement whole genome amplification", Nucleic Acids Research, 32(9), 2004, 1-11.

Page, S. L. et al., "Chromosome Choreography: The Meiotic Ballet", Science, 301, 2003, 785-789.

Paik, P. K. et al., "Next-Generation Sequencing of Stage IV Squamous

Paik, P. K. et al., "Next-Generation Sequencing of Stage IV Squamous Cell Lung Cancers Reveals an Association of P13K Aberrations and Evidence of Clonal Heterogeneity in Patients with Brain Metastases", Cancer Discovery, vol. 5, Apr. 30, 2015, 610-621.

Pakstis, et al., "Candidate SNPs for a Universal Individual Identification Panel", Human Genetics, vol. 121, No. 3-4,, Feb. 27, 2007, 305-317.

Pakstis, et al., "SNPS for Individual Identification", Forensic Science International, vol. 1, May 22, 2008, 479-481.

Palka-Santini, Maria et al., "Large Scale Multiplex PCR Improves Pathogen Detection by DNA Microarrays", BMC Microbiology, vol. 9, No. 1, 2009, 14 pages.

Palomaki, G. E. et al., "DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study", Genetics in Medicine, vol. 13, No. 1, Nov. 2011, 913-920.

Palomaki, Glenn et al., "DNA Sequencing of Maternal Plasma Reliably Identifies Trisomy 18 and Trisomy 13 as Well as Down Syndrome: an International Collaborative Study", Genetics in Medicine, 2012, 10.

Palomaki, Glenn E. et al., "DNA Sequencing of Maternal Plasma to Detect Down Syndrome: An International Clinical Validation Study", Genetics in Medicine (pre-print version), 13, 2011, 8 pgs.

Panjkovich, Alejandro et al., "Comparison of Different Melting Temperature Calculation Methods for Short DNA Sequences", Bioinformatics, vol. 21, 2005, pp. 711-722.

Bioinformatics, vol. 21, 2005, pp. 711-722. Papadopoulou, E. et al., "Cell-Free DNA and RNA in Plasma as a New Molecular Marker for Prostate Cancer", Oncology Research, vol. 14, 2004, 439-445.

Papageorgiou, Elisavet A. et al., "Fetal-Specific DNA Methylation Ratio Permits Noninvasive Prenatal Diagnosis of Trisomy 21", Nature Medicine (advance online publication), 17, 2011, 5 pgs.

Nature Medicine (advance online publication), 17, 2011, 5 pgs. Parameswaran, P. et al., "A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing", NucleicAcids Research, vol. 35, No. 19, Oct. 11, 2007, 9 pages.

Park, et al., "First-Line Erlotinib Therapy Until and Beyond Response Evaluation Criteria in Solid Tumors Progression in Asian Patients With Epidermal Growth Factor Receptor Mutation-Positive Non-Small-Cell Lung Cancer", JAMA Oncol., 2(3), 2015, 305-312.

Parker, A. V. et al., "The Effect of Sodium Citrate on the Stimulation of Polymorphonuclear Leukocytes", Investigative Ophthalmology & Visual Science, vol. 26, 1985, 1257-1261.

Paruzynski, A. et al., "Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing", Nature Protocols, vol. 5, No. 8, Jul. 8, 2010, 1379-1395.

Pask, R. et al., "Investigating the utility of combining 29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray TM genotyping", BMC Biotechnology, vol. 4, No. 15, Jul. 27, 2004, 8 pages.

Pastinen, T. et al., "Minisequencing: A Specific Tool for DNA Analysis and Diagnostics on Oligonucleotide Arrays", Genome Research, vol. 7, 1997, 606-614.

Pathak, A. et al., "Circulating Cell-Free DNA in Plasma/Serum of Lung Cancer Patients as a Potential Screening and Prognostic Tool", Clinical Chemistry, 52, 2006, 1833-1842.

Patil, N. et al., "Blocks of Limited Haplotype Diversity Revealed by High-Resolution Scanning of Human Chromosome 21", Science, vol. 294, Nov. 23, 2001, 1719-1723.

vol. 294, Nov. 23, 2001, 1719-1723. Paunio, T. et al., "Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA", Clinical Chemistry, vol. 42, No. 9, 1996, 1382-1390.

PCT/US2006/045281, "International Preliminary Reporton Patentability", dated May 27, 2008, 1 pg. PCT/US2006/045281, "International Search Report and Written

PCT/US2006/045281, "International Search Report and Written Opinion", dated Sep. 28, 2007, 7 pgs.
PCT/US2008/003547, "International Search Report", dated Apr. 15,

PCT/US2008/003547, "International Search Report", dated Apr. 15, 2009, 5 pgs.

PCT/US2009/034506, "International Search Report", dated Jul. 8, 2009, 2 pgs.

PCT/US2009/045335, "International Search Report", dated Jul. 27, 2009, 1 pg.

PCT/US2009/052730, "International Search Report", dated Sep. 28, 2009, 1 pg.
PCT/US2010/050824, "International Search Report", dated Nov.

15, 2010, 2 pgs. PCT/US2011/037018, "International Search Report", dated Sep. 27,

2011, 2 pgs. PCT/US2011/061506, "International Search Report", dated Mar.

PC1/US2011/061506, "International Search Report", dated Mar. 16, 2012, 1 pgs.

PCT/US2011/066938, "International Search Report", dated Jun. 20, 2012, 1 pg.

## US 11,530,454 B2

Page 23

### (56) References Cited

#### OTHER PUBLICATIONS

PCT/US2012066339, "International Search Report", dated Mar. 5, 2013. 1 pg.

PCT/US2013/028378, "International Search Report and Written Opinion", dated May 28, 2013, 11 pgs.

PCT/US2013/57924, "International Search Report and Written Opinion", dated Feb. 18, 2014, 8 pgs.

PCT/US2014/051926, "International Search Report and Written Opinion", dated Dec. 9, 2014, 3 pgs.

Pearson, K., "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", Philosophical Magazine Series 5, vol. 50, Issue 302, 1900. 157-175.

Pelizzari, C. A. et al., "Quantitative analysis of DNA array autoradiographs", Nucleic Acids Research, vol. 28, No. 22, 2000, 4577-4581

Pena, Sergio D.J et al., "Paternity Testing in the DNA Era", Trends in Genetics, 10, 6, 1994, 204-209.

Perakis, S. et al., "Advances in Circulating Tumor DNA Analysis", Advances in Clinical Chemistry, vol. 80, 2017, 73-153.

Pergament, E. et al., "Single-Nucleotide Polymorphism-Based Noninvasive Prenatal Screening in a High-Risk and Low-Risk Cohort", Obstetrics & Gynecology, vol. 124, No. 2, Part 1, Aug. 2014, 210-218 + Appendices.

Perkel, Jeffrey M., "Overcoming the Challenges of Multiplex PCR", Biocompare Editorial Article, 2012, 1-5.

Perry, George H. et al., "The Fine-Scale and Complex Architecture of Human Copy-Number Variation", The American Journal of Human Genetics, 82, 2008, 685-695.

Pertl, B. et al., "Detection of Male and Female Fetal DNA in Maternal Plasma by Multiplex Fluorescent Polymerase Chain Reaction Amplification of Short Tandem Repeats", Hum. Genet., 106, 2000, 45-49.

Peters, D., "List of Materials Considered by David Peters, Ph.D.", Jun. 13, 2019, 2 pages.

Peters, David P. et al., "Noninvasive Prenatal Diagnosis of a Fetal Microdeletion Syndrome", New England Journal of Medicine, 365(19), 2011, 1847-1848.

Pfaffl, Michael W., "Quantification Strategies in Real-time PCR", A-Z of quantitative PCR, 2004, pp 87-112.

Pfaffl, Michael W., "Relative Expression Software Tool (REST©) for Group-Wise Comparison and Statistical Analysis of Relative Expression Results in real-Time PCR", Nucleic Acids Research, 30(9), 2002, 10 pgs.

Philip, J. et al., "Late First-Trimester Invasive Prenatal Diagnosis Results of an International Randomized Trial", American College of Obstetricians and Gynecologists, vol. 103, No. 6, Jun. 2004, 1164-1173.

Phillips, C. et al., "Resolving Relationship Tests that Show Ambiguous STR Results Using Autosomal SNPs as Supplementary Markers", Forensic Science International: Genetics 2, 2008, 198-204.

Pinard, et al., "Assessment of Whole Genome Amplification-induced Bias Through High-throughput, Massively Parallel Whole Genome Sequencing", BMC Genomics, vol. 7:216, Aug. 23, 2006, 1-21.

Pirker, C. et al., "Whole Genome Amplification for CGH Analysis Linker-Adapter PCR as the Method of Choice for Difficult and Limited Samples", Cytometry Part A, vol. 61 A, 2004, 26-34.

Podder, Mohua et al., "Robust SN P genotyping by multiplex PCR and arrayed primer", BMC Medical Genomics, I(5), 2008, 1-15. Poirier, K. et al., "Maternal mosaicism for mutations in the ARX

Poirier, K. et al., "Maternal mosaicism for mutations in the ARX gene in a family with X linked mental retardation", Human Genetics, vol. 118, Aug. 3, 2005, 45-48.

Pont-Kingdon, G. et al., "Rapid Detection of Aneuploidy (Trisomy 21) by Allele Quantification Combined with Melting Curves Analysis of Single-Nucleotide Polymorphism Loci", Clinical Chemistry, vol. 49, No. 7, 2003, 1087-1094.

Poon, L. L. et al., "Differential DNA Methylation between Fetus and Mother as a Strategy for Detecting Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 48, No. 1, 2002, 35-41. Popova, T. et al., "Genome Alteration Print (GAP): a tool to

Popova, T. et al., "Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays", Genome Biology, vol. 10, R128, Nov. 11, 2009, 1-14. Porreca, Gregory J. et al., "Multiplex Amplification of Large Sets of Human Exons", Nature Methods, 4, (advance online publication), 2007 6

Pourmand, et al., "Multiplex Pyrosequencing", Nucleic Acid Research, vol. 30, No. 7, Apr. 1, 2002, 1-5.

Prabhu, et al., "Overlapping Pools for High-throughput Targeted Resequencing", Genome Research, vol. 19, May 15, 2009, 1254-1261.

Price, T.S. et al., ""SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data",", Nucleic Acids Research, vol. 33, No. 11, Jun. 16, 2005, 3455-3464.

Primdahl, H. et al., "Allelic Imbalances in Human Bladder Cancer: Genome-Wide Detection With High-Density Single-Nucleotide Polymorphism Arrays", Journal of the National Cancer Institute, vol. 94, No. 3, Feb. 6, 2002, 216-223.

Profitt, J et al., "Isolation and Characterisation of Recombination Events Involving Immunoglobulin Heavy Chain Switch Regions in Multiple Myeloma Using Long Distance Vectorette PCR (Ldv-pcr)", Leukemia, vol. 13, No. 7, Jul. 1999, 1100-1107. Puszyk, William M. et al., "Noninvasive Prenatal Diagnosis of

Puszyk, William M. et al., "Noninvasive Prenatal Diagnosis of Aneuploidy Using Cell-free Nucleic Acids in Maternal Blood: Promises and Unanswered Questions", Prenatal Diagnosis, vol. 28, No. 1, Nov. 16, 2007, 1-6.

Qiagen, QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook, Feb. 2003 ("Qiagen (2003)"), 2003, 68 pages.

Qin, Z. S. et al., "Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms", Am. J. Hum Genet., vol. 71, 2002, 1242-1247. Quan, P. C. et al., "Studies on the mechanism of NK cell lysis", The Journal of Immunology, vol. 128, 1982, 1786-1791.

Quinlan, M. P., "Amniocentesis: Indications and Risks", American Medical Association Journal of Ethics: Virtual Mentor, vol. 10, No. 5, May 2008, 304-306.

Quinn, G. P. et al., "Experimental Design and Data Analysis for Biologists", Graphical Exploration of Data, 2002, 64-67.

Rabinowitz, et al., "Accurate Prediction of HIV-1 Drug Response from the Reverse Transcriptase and Protease Amino Acid Sequences Using Sparse Models Created by Convex Optimization", Bioinformatics, 22, 5, 2006, 541 -549.

Rabinowitz, M., "A System and Method for Integrating, Validating and Applying Genetic and Clinical Data to Enhance Medical Decisions", Nov. 29, 2005, 155 pgs.

Decisions", Nov. 29, 2005, 155 pgs. Rabinowitz, Matthew et al., "Origins and rates of aneuploidy inhuman blastomeres", Fertility and Sterility, vol. 97, No. 2, Feb. 2012, 395-401.

Rabinowitz, Matthew et al., "Non-Invasive Prenatal Aneuploidy Testing of Chromosomes 13, 18, 21, X, and Y Using Targeted Sequencing of Polymorphic Loci", The American Society of Human Genetics, meeting poster, 2012, 1 pg.
Rachlin, J. et al., "Computational tradeoffs in multiplex PCR assay

Rachlin, J. et al., "Computational tradeoffs in multiplex PCR assay design for SNP genotyping", BMC Genomics, vol. 6, No. 102, Jul. 25, 2005, 11 pages.

Raemdonck, Dirk Van et al., "Ex-vivo lung perfusion", Transplant International, vol. 28, Issue 6, Special Issue: Focus Issue: Machine Perfusion, 2014, 643-656.

Ragoussis, J., "Genotyping Technologies for Genetic Research", Annual Review of Genomics and Human Genetics, vol. 10 (1), Sep. 1, 2009, 117-133.

Rahmann, Sven et al., "Mean and variance of the Gibbs free energy of oligonucleotides in the nearest neighbor model under varying conditions", Bioinformatics, 20(17), 2004, 2928-2933.

Raindance Technologies, "Multiplexing with RainDrop Digital PCR", Application Note, 2013, 2 pgs.

Raindance Technologies, et al., "RainDance Technologies Introduces the RDT 1000", RainDance Technologies, Nov. 12, 2008.

## US 11,530,454 B2

Page 24

### (56) References Cited

#### OTHER PUBLICATIONS

Rava, Richard P. et al., "Circulating Fetal Cell-Free DNA Fraction Differ in Autosomal Aneuploidies and Monosomy X", Clinical Chemistry, 60(1), (papers in press), 2013, 8 pgs.

Ravipati, Goutham et al., "Comparison of Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of Stress Testing Versus 64-Multislice Coronary Computed Tomography Angiography in Predicting Obstructive Coronary Artery Disease Diagnosed by Coronary Angiogr", The American Journal of Cardiology, Coronary Artery Disease. vol. 101, Issue 6, Mar. 15, 2008, 774-775

Rechitsky, S. et al., "Allele Dropout in Polar Bodies and Blastomeres", Journal of Assisted Reproduction and Genetics, vol. 15, No. 5, 1998, 253-257

Rechitsky, Svetlana et al., "Preimplantation Genetic Diagnosis with HLA Matching", Reproductive Bio Medicine Online, 9, 2, 2004, 210-221.

Reeves, R. H. et al., "Too much of a good thing: mechanisms of gene action in Down syndrome", Trends in Genetics, vol. 17, No. 2, Feb. 2, 2001, 83-88.

Reinert, T. et al., "Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery", Gut, vol. 65, 2016, 625-634.

Renwick, P. et al., "Proof of Principle and First Cases Using Preimplantation Genetic Haplotyping—A Paradigm Shift for Embryo Diagnosis", Reproductive BioMedicine Online, 13 (1), 2006, 110-119

Rhoads, A. et al., "PacBio Sequencing and Its Applications", Genomics Proteomics Bioinformatics, vol. 13, Nov. 2, 2015, 278-289

Ricciotti, Hope, "Eating by Trimester", Online]. Retrieved from Internet:<a href="http://www.youandyourfamily.com/article.php?story=Eating+by+Trimester">http://www.youandyourfamily.com/article.php?story=Eating+by+Trimester</a>, 2014, 3.

Riley, D. E., "DNA Testing: An Introduction for Non-Scientists an Illustrated Explanation", Scientific Testimony: An Online Journal, http://www.scientific.org/tutorials/articles/riley/riley.html, Apr. 6, 2005, 22, pages.

Riva, F., "Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer", Clinical Chemistry, vol. 63, No. 3, 2017, 691-699.

Robertson, G. et al., "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing", Nature Methods, vol. 4, No. 8, Aug. 2007, 651-657. Roche Diagnostics, et al., "Versatile Nucleic Acid Purification", MagnaPure Manual, Feb. 3, 2012. Rogaeva, E. et al., "The Solved and Unsolved Mysteries of the

Rogaeva, E. et al., "The Solved and Unsolved Mysteries of the Genetics of Early-Onset Alzheimer's Disease", NeuroMolecular Medicine, vol. 2, 2002, 1-10.

Roman, B. L. et al., "Non-Radioisotopic AFLP Method Using PCR Primers Fluorescently Labeled with CyA 5", BioTechniques, vol. 26, Feb. 1999, 236-238.

Roper, Stephen M. et al., "Forensic Aspects of DNA-Based Human Identity Testing", Journal of Forensic Nursing, 4, 2008, 150-156. Rosado, J. A. et al., "Tyrosine kinases activate store-mediated Ca2+entry in human platelets through the reorganization of the actin cytoskeleton", Biochem. J., vol. 351, 2000, 429-437.

Rosen, D. R. et al., "Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis", Nature, vol. 362, Mar. 4, 1993, 59-62.

Ross, P. et al., "Quantitative Approach to Single-Nucleotide Polymorphism Analysis Using MALDI-TOF Mass Spectrometry", BioTechniques, vol. 29, Sep. 2000, 620-629.

Rothberg, et al., "The Development and Impact of 454 Sequencing", Nature Biotechnology, vol. 26, No. 10, Oct. 9, 2008, 1117-1124. Roux, K., "Optimization and Troubleshooting in PCR", PCR Methods Appl. 4, 1995, 185-194.

Rozen, Steve et al., "Primer3 on the WWW for General Users and for Biologis Programmers", Methods in Molecular Biology, 132: Bioinformatics Methods and Protocols, 1999, 365-386.

Ruano, G. et al., "Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules", Proc. Natl. Acad. Sci. USA, vol. 87, Aug. 1990, 6296-6300.

Rubio, J. M. et al., "Semi-nested, Multiplex Polymerase Chain Reaction for Detection of Human Malaria Parasites and Evidence of Plasmodium Vivax Infection in Equatorial Guinea", The American Journal of Tropical Medicine and Hygiene, vol. 60, 1999, pp. 183-187.

Ruschendorf, et al., "Alohomora: A Tool For Linkage Analysis Using 10K SNP Array Data", Bioinformatics Applications Notes, vol. 21, No. 9, Jan. 12, 2005, 2123-2125.

Russell, L. M., "X Chromosome Loss and Ageing", Cytogenetic and Genome Res., 116, 2007, 181-185.

Ryan, A. et al., "Informatics-Based, Highly Accurate, Noninvasive Prenatal Paternity Testing", Genetics in Medicine (advance online publication), 2012, 5 pgs.

Ryan, B. M. et al., "A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up", Gut, vol. 52, 2003, 101-108.

Rychlik, et al., "Optimization of the annealing temperature for DNA amplification in vitro", Nucleic Acids Research, 18(21), 1990, 6409-6412.

Sahota, A., "Evaluation of Seven PCR-Based Assays for the Analysis of Microchimerism", Clinical Biochemistry, vol. 31, No. 8., 1998, 641-645.

Sahukhal, G. S. et al., "msaABCR operon positively regulates biofilm development by repressing proteases and autolysis in *Staphlococcus aureus*", FEMS Microbiology Letters, vol. 362, No. 4, 2015, 1-10.

Saito, H. et al., "Prenatal DNA diagnosis of a single-gene disorder from maternal plasma", The Lancet, vol. 356, Sep. 30, 2000, 1170. Saker, A. et al., "Genetic characterisation of circulating fetal cells allows non-invasive prenatal diagnosis of cystic fibrosis", Prenatal Diagnosis, vol. 26, Jul. 11, 2006, 906-916.

Samango-Sprouse, C. et al., "SNP-Based Non-Invasive Prenatal Testing Detects Sex Chromosome Aneuploidies with High Accuracy", Prenatal Diagnosis, 33, 2013, 1-7.

Samura, O. et al., "Diagnosis of Trisomy 21 in Fetal Nucleated Erythrocytes from Maternal Blood by Use of Short Tandem Repeat Sequences", Clinical Chemistry, vol. 47, No. 9, 2001, 1622-1626. Sanchez, C. et al., "New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis", Nature Partner Journals, vol. 3, No. 31, Nov. 23, 2018, 12 pgs.

Sander, Chris, "Genetic Medicine and the Future of Health Care", Science, 287(5460), 2000, 1977-1978.

Sanger, et al., "Nucleotide Sequence of Bacteriophage Lambda DNA", Journal of Molecular Biology, vol. 162, No. 4, Dec. 25, 1982, 729-773.

Santalucia, J. et al., "The Thermodynamics of DNA Structural Motifs", Annu. Rev. Biophys. Biomol. Struct., 33, 2004, 415-440. Santalucia, John J.R. et al., "Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability", Biochemistry, 35, 1996, 3555-3562.

Santalucia, Jr., J., "Physical Principles and Visual-OMP Software for Optimal PCR Design", Methods in Molecular Biology, vol. 402, 2007. 3-33.

Sasabe, Yutaka , "Genetic Diagnosis of Gametes and Embryos Resulting from ART", Japanese Journal of Fertility and Sterility, vol. 46, No. 1, 2001, 43-46.

Scarpa, A. et al., "Molecular Typing of Lung Adenocarcinoma on Cytological Samples Using a Multigene Next Generation Sequencing Panel", PLOS One, vol. 8, No. 11, Nov. 13, 2013, 6 pgs.

Schaaf, C. P. et al., "Copy Number and SNP Arrays in Clinical Diagnostics", Annu. Rev. Genomics Hum. Genet., vol. 12, 2011, 25-51.

Scheet, P. et al., "A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase", The American Journal of Human Genetics, vol. 78, Apr. 2006, 629-644.

Case: 24-1324 Document: 42-1 Page: 339 Filed: 03/18/2024

## US 11,530,454 B2

Page 25

#### (56)References Cited

#### OTHER PUBLICATIONS

Schmitt, M. W. et al., "Detection of ultra-rare mutations by nextgeneration sequencing", PNAS, vol. 109, No. 36, Sep. 4, 2012, 14508-

Schoske, R et al., "Multiplex PCR Design Strategy used for the Simultaneous Amplification of 10 Y Chromosome Short Tandem Repeat (STR) Loci", Analytical and Bioanalytical Chemistry, vol. 375, 2003, 333-343.

Schoumans, J et al., "Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH)", JMed Genet, 42, 2005, 699-705.

Schubert, "Picking out prenatal DNA", Nature Medicine, vol. 10, No. 785, Aug. 2004, 1 page.

Schutz, E. et al., "Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: A prospective, observational, multicenter cohort study", PLOS Medicine, vol. 14, No. 4, Apr. 25, 2017, 19 pgs. Schwarzenbach, H. et al., "Cell-free nucleic acids as biomarkers in

cancer patients", Nature Reviews: Cancer, vol. 11, Jun. 2011, 426-437

Schwarzenbach, H. et al., "Detection and Characterization of Circulating Microsatellite-DNA in Blood of Patients with Breast Cancer", Ann. N.Y. Acad. Sci., vol. 1022, 2004, 25-32.

Schwarzenbach, H. et al., "Evaluation of cell-free tumour DNA and RNA in patients with breast cancer and benign breast disease",

Molecular BioSystems, vol. 7, 2011, 2848-2854. Sebat, Jonathan et al., "Strong Association of De Novo Copy Number Mutations with Autism", Science, 316, 2007, 445-449.

Sehnert, A. et al., "Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood", Clinical Chemistry (papers in press), 57 (7), 2011, 8 pgs.

Selzner, Markus et al., "Normothermic Ex Vivo Liver Perfusion Using Steen Solution as Perfusate for Human Liver Transplantation: First North American Results", Liver Transplantation, vol. 22, Issue

Seppo, A. et al., "Detection of circulating fetal cells utilizing automated microscopy: potential for noninvasive prenatal diagnosis of chromosomal aneuploidies", Prenatal Diagnosis, vol. 28, Jul. 22, 2008, 815-821.

Sermon, Karen et al., "Preimplantation genetic diagnosis", The Lancet, Lancet Limited. 363(9421), 2000, 1633-1641

Servin, B et al., "MOM: A Program to Compute Fully Informative Genotype Frequencies in Complex Breeding Schemes", Journal of Heredity, vol. 93, No. 3, Jan. 1, 2002 (Jan. 1, 2002), pp. 227-228. Sethi, Himanshu et al., "Analytical validation of the Signatera (TM) Ruo assay, a highly sensitive patient-specific multiplex PCR NGSbased noninvasive cancer recurrence detection and therapy monitoring assay", Cancer Research, vol. 78, No. 13, 2018, 4542.

Sham, P. et al., "DNA Pooling: A Tool for Large-Scale Association

Studies", Nature Reviews Genetics, vol. 3, Nov. 2002, 862-871. Shapero, M. H. et al., "MARA: A Novel Approach for Highly Multiplexed Locus-specific SNP Genotyping Using High-density DNA Oligonucleotide Arrays", Nucleic Acids Research, vol. 32, No. 22, 2004, 1-9.

Sharples, et al., "Diagnostic Accuracy of Coronary Angiography and Risk Factors for Post-heart-transplant Cardiac Allograft Vasculopathy", Transplantation, vol. 76, No. 4, Aug. 27, 2003,

Shaw-Smith, et al., "Microarray Based Comparative Genomic Hybridisation (array-CGH) Detects Submicroscopic Chromosomal Deletions and Duplications in Patients with Learning Disability/ Mental Retardation and Dysmorphic Features", J. Med. Genet., 41,

Shen, et al., "High-quality DNA sequence capture of 524 disease candidate genes", High-quality DNA sequence capture of 524 disease candidate genes, Proceedings of the National Academy of Sciences, vol. 108, No. 16, Apr. 5, 2011 (Apr. 5, 2011), pp. 6549-6554.

Shen, R. et al., "High-throughput SNP genotyping on universal bead arrays", Mutation Research, vol. 573, Feb. 11, 2005, 70-82.

Shen, Zhiyong, "MPprimer: a program for reliable multiplex PCR primer design", BMC Bioinformatics 2010, 11:143, 1-7.

Shendure, J. et al., "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome", Science, Nov. 30, 2007, 18-24.

Shendure, J. et al., "Next-generation DNA sequencing", Nature Biotechnology, vol. 26, No. 10, Oct. 2008, 1135-1145.

Sherlock, J et al., "Assessment of Diagnostic Quantitative Fluorescent Multiplex Polymerase Chain Reaction Assays Performed on Single Cells", Annals of Human Genetics, 62, 1, 1998, 9-23

Shi, H. et al., "Melanoma whole-exome sequencing identifies V600E B-RAF amplification-mediated acquired B-RAF inhibitor resistance", Nature Communications, vol. 3, No. 724, Mar. 6, 2012,

Shinozaki, M. et al., "Utility of Circulating B-RAF DNA Mutation in Serum for Monitoring Melanoma Patients Receiving Biochemotherapy", Clin Cancer Res, vol. 13, No. 7, Apr. 1, 2007, 2068-2074

Shiroguchi, K. et al., "Digital RNA sequencing minimizes sequencedependent bias and amplification noise with optimized singlemolecule barcodes", PNAS, vol. 109, No. 4, Jan. 24, 2012, 1347-1352.

Shokralla, S. et al., "Next-generation DNA barcoding: using nextgeneration sequencing to enhance and accelerate DNA barcode capture from single specimens". Molecular Ecology Resources, vol. 14. 2014, 892-901.

Short, N. J. et al., "Targeted next-generation sequencing of circulating cell-free DNA vs bone marrow in patients with acute myeloid leukemia", Blood Advances, vol. 4, No. 8, Apr. 23, 2020, 1670-

Shyamala, Venkatakrishna et al., "Genome Walking by Single-Specific-Primer Polymerase Chain Reaction: SSP-PCR", Gene, vol. 84, 1989, pp. 1-8.

Siebert, P. D. et al., "An improved PCR method for walking in uncloned genomic DNA", Nucleic Acids Research, vol. 23, No. 6, 1995, 1087-1088.

Sigdel, T. et al., "Plasma Donor-Derived Cell-Free DNA Quantification by massively multiplex PCR Distinguishes Kidney Transplant Acute Rejection", Transplantation, vol. 102, No. 7S, Jul. 2018, S178-S179

Sigdel, T. K. et al., "Optimizing Detection of Kidney Transplant Injury by Assessment of Donor-Derived Cell-Free DNA via Massively Multiplex PCR", Journal of Clinical Medicine, vol. 8, No. 19,

Dec. 23, 2018, 17 pages. Simpson, J. et al., "Fetal Cells in Maternal Blood: Overview and Historical Perspective", Annals New York Academy of Sciences, 731, 1994, 1-8

Singh, Vinayak K. et al., "PCR Primer Design", Molecular Biology

Today, vol. 2, 2001, pp. 27-32. Sint, Daniela et al., "Advances in Multiplex PCR: Balancing Primer Efficiencies and Improving Detection Success", Methods in Ecology and Evolution, 3, 2012, 898-905. Sivertsson, A. et al., "Pyrosequencing as an Alternative to Single-

Strand Conformation Polymorphism Analysis for Detection of N-ras Mutations in Human Melanoma Metastases", Clinical Chemistry, vol. 48, No. 12, 2002, 2164-2170. Slater, Howard et al., "High-Resolution Identification of Chromo-

somal Abnormalities Using Oligonucleotide Arrays Containing 116,204 SNPs", Am. J. Hum. Genet., 77, 5, 2005, 709-726.

Smith, et al., "Rapid Whole-genome Mutational Profiling using Next-generation Sequencing Technologies", Genome Research, vol. 18, Sep. 4, 2008, 1638-1642.

Smith, James F. et al., "Cell-free Fetal DNA in Maternal Plasma", Neo Reviews, vol. 9, No. 8, Aug. 1, 2008, e332-e337

Snijders, Antoine et al., "Assembly of Microarrays for Genome-Wide Measurement of DNA Copy Number", Nature Genetic, 29, 2001, 263-264.

Snyder, T. M. et al., "Universal noninvasive detection of solid organ transplant rejection", PNAS, vol. 108, No. 15, Apr. 12, 2011, 6229-6234

Societies Related to Genetic Med, "Guideline related to genetic examination", Japanese Society for Genetic Counseling, Japanese

# US 11,530,454 B2

Page 26

#### (56) References Cited

#### OTHER PUBLICATIONS

Society for Gene Diagnosis and Therapy, Japan Society of Obstetrics and Gynecology, 2003, 2-15.

Solexa, "Application Note: DNA Sequencing", 2006, pp. 1-2.

Solomon, M. J. et al., "Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures", Proc. Natl. Acad. Sci. USA, vol. 82, 1985, 6470-6474.

Sorenson, G. D. et al., "Soluble Normal and Mutated DNA Sequences from Single-Copy Genes in Human Blood", Cancer Epdemiology, Biomarkers & Prevention, vol. 3, Jan./Feb. 1994, 67-71.

sourceforge.net, "Primer3", http://primer3.sourceforge.net/, 2009, 1 pg.

Sparks, A. et al., "Non-Invasive Prenatal Detection and Selective Analysis of Cell-Free DNA Obtained from Maternal Blood: Evaluation for Trisomy 21 and Trisomy 18", American Journal of Obstetrics & Gynecology 206, 2012, 319.e1-319.e9.

Sparks, Andrew B. et al., "Selective Analysis of Cell-Free DNA in

Sparks, Andrew B. et al., "Selective Analysis of Cell-Free DNA in Maternal Blood for Evaluation of Fetal Trisomy", Prenatal Diagnosis, 32, 2012, 1-7.

Spencer, K. et al., "Maternal serum levels of dimeric inhibin A in pregnancies affected by trisomy 21 in the first trimester", Prenatal Diagnosis, vol. 21, 2001, 441-444.

Spencer, K. et al., "Maternal serum levels of total activin-A in first-trimester trisomy 21 pregnancies", Prenatal Diagnosis, vol. 21, 2001, 270-273.

Spertini, D. et al., "Screening of Transgenic Plants by Amplification of Unknown Genomic DNA Flanking T-DNA", BioTechniques, vol. 27, Aug. 1999, 308-314.

Spes, et al., "Diagnostic And Prognostic Value of Serial Dobutamine Stress Echocardiography for Noninvasive Assessment of Cardiac Allograft Vasculopathy: A Comparison With Coronary Angiography and Intravascular Ultrasound", Circulation, vol. 100, No. 5, Aug. 3, 1999, 509-515.

Spindler, K.L. G. et al., "Cell-free DNA in healthy individuals, noncancerous disease and strong prognostic value in colorectal cancer", International Journal of Cancer, vol. 135, 2014, 2984-2991. Spindler, K.-L. G. et al., "Cell-Free DNA in Metastatic Colorectal Cancer: A Systematic Review and Meta-Analysis", The Oncologist, vol. 22, 2017, 1049-1055.

Spiro, Alexander et al., "A Bead-Based Method for Multiplexed Identification and Quantitation of DNA Sequences Using Flow Cytometry", Applied and Environmental Microbiology, 66, 10, 2000, 4258-4265.

Spits, C et al., "Optimization and Evaluation of Single-Cell Whole Genome Multiple Displacement Amplification", Human Mutation, 27(5), 496-503, 2006.

Srinivasan, et al., "Noninvasive Detection of Fetal Subchromosome Abnormalities via Deep Sequencing of Maternal Plasma", The American Journal of Human Genetics 92, 167-176, Feb. 7, 2013. Stephens, M. et al., "Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation", Am. J. Hum. Genet., vol. 76, 2005, 449-462.

Stephens, Mathews, et al., "A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data", Am. J. Hum. Genet., 73, 2003, 1162-1169.

Stevens, Robert et al., "Ontology-Based Knowledge Representation for Bioinformatics", Briefings in Bioinformatics, 1, 4, 2000, 398-414.

Stewart, C. M. et al., "Circulating cell-free DNA for non-invasive cancer management", Cancer Genetics, vol. 228-229, 2018, 169-179

Stewart, S. et al., "Revision of the 1990 Working Formulation for the Standardization of Nomenclature in the Diagnosis of Heart Rejection", The Journal of Heart and Lung Transplantation, vol. 24, No. 11, Nov. 2005, 1710-1720.

Steyerberg, E.W et al., "Application of Shrinkage Techniques in Logistic Regression Analysis: A Case Study", Statistica Neerlandica, 55(1), 2001, 76-88.

Stiller, et al., "Direct Multiplex Sequencing (DMPS)—A Novel Method for Targeted High-thoroughput Sequencing of Ancient and Highly Degraded DNA", Genome Research, vol. 19, No. 10, Jul. 27, 2009, 1843-1848.

Stolerman, Elliot S. et al., "Haplotype structure of the ENPP1 Gene and Nominal Association of the K121Q missense single nucleotide polymorphism with glycemic traits in the Framingham Heart Study", Diabetes. vol. 57. Issue 7. Jul. 1. 2008. 1971-1977.

Diabetes, vol. 57, Issue 7, Jul. 1, 2008, 1971-1977.
Stone, J. P. et al., "Ex Vivo Normothermic Perfusion Induces Donor-Derived Leukocyte Mobilization and Removal Prior to Renal Transplantation", Kidney Int Rep., vol. 1, No. 4, Aug. 6, 2016, 230-239.

Strom, C. et al., "Three births after preimplantation genetic diagnosis for cystic fibrosis with sequential first and second polar body analysis", American Journal of Obstetrics and Gynecology, 178 (6), 1998, 1298-1306.

Strom, Charles M. et al., "Neonatal Outcome of Preimplantation Genetic Diagnosis by Polar Body Removal: The First 109 Infants", Pediatrics, 106(4), 2000, 650-653.

Stroun, Maurice et al., "Prehistory of the Notion of Circulating Nucleic Acids in Plasma/Serum (CNAPS): Birth of a Hypothesis", Ann. N.Y. Acad. Sci., 1075, 2006, 10-20.
Su, S.Y. et al., ""Inferring combined CNV/SNP haplotypes from

Su, S.Y. et al., ""Inferring combined CNV/SNP haplotypes from genotype data"", Bioinformatics, vol. 26, No. 11,1, Jun. 1, 2010, 1437-1445.

Su, Z. et al., "A Platform for Rapid Detection of Multiple Oncogenic Mutations With Relevance to Targeted Therapy in Non-Small-Cell Lung Cancer", The Journal of Molecular Diagnostics,, vol. 13, No. 1, Jan. 2011, 74-84.

Sun, Guihua et al., "SNPs in human miRNA genes affect biogenesis and function", RNA, 15(9), 2009, 1640-1651.

Swarup, V. et al., "Circulating (cell-free) nucleic acids—A promising, non-invasive tool for early detection of several human diseases", FEBS Letters, vol. 581, 2007, 795-799.

Sweet-Kind Singer, J. A. et al., "Log-penalized linear regression", IEEE International Symposium on Information Theory, 2003. Proceedings, 2003, 286.

Swinkels, D. W. et al., "Effects of Blood-Processing Protocols on Cell-free DNA Quantification in Plasma", Clinical Chemistry, vol. 49, No. 3, 2003, 525-526.

49, No. 3, 2003, 525-526. Syvanen, A.C. "Toward genome-wide SNP genotyping", Nature Genetics Supplement, vol. 37, Jun. 2005, S5-S10.

Taback, B. et al., "Quantification of Circulating DNA in the Plasma and Serum of Cancer Patients", Ann. N.Y. Acad. Sci, vol. 1022, 2004. 17-24.

Takala, et al., "A High-throughput Method for Quantifying Alleles and Haplotypes of the Malaria Vaccine Candidate Plasmodium Falciparum Merozoite Surface Protein-1 19 kDa", Malaria Journal, vol. 5:31, Apr. 20, 2006, 1-10.

Takano, T. et al., "Epidermal Growth Factor Receptor Gene Muta-

Takano, T. et al., "Epidermal Growth Factor Receptor Gene Mutations and Increased Copy Numbers Predict Gefitinib Sensitivity in Patients With Recurrent Non-Small-Cell Lung Cancer", Journal of Clinical Oncology, vol. 23, No. 28, Oct. 1, 2005, 6829-6837.

Takara Biomedicals, "Competitive PCR Guide", Lit. # L0126, Aug. 1999, 9 pages.

Takashima, Y. et al., "Expansion-contraction of photoresponsive artificial muscle regulated by host-guest interactions", Nature Communications, vol. 3, No. 1270, Dec. 11, 2012, 8 pages.

Taliun, D. et al., "Efficient haplotype block recognition of very long and dense genetic sequences", BMC Bioinformatics, vol. 15 (10), 2014, 1-18.

Tamura, et al., "Sibling Incest and formulation of paternity probability case report", Legal Medicine, 2000, vol. 2, p. 189-196.

Tang, et al., Multiplex fluorescent PCR for noninvasive prenatal detection of fetal-derived paternally inherited diseases using circulatory fetal DNA in maternal plasma, Eur J Obstet Gynecol Reprod Biol, 2009, v.144, No. 1, p. 35-39.

Tang, N. et al., "Detection of Fetal-Derived Paternally Inherited X-Chromosome Polymorphisms in Maternal Plasma", Clinical Chemistry, 45 (11), 1999, 2033-2035.

istry, 45 (11), 1999, 2033-2035. Tebbutt, S. J. et al., "Microarray genotyping resource to determine population stratification in genetic association studies of complex disease", BioTechniques, vol. 37, Dec. 2004, 977-985.

# US 11,530,454 B2

Page 27

### (56) References Cited

#### OTHER PUBLICATIONS

Ten Bosch, J., "Keeping Up With the Next Generation Massively Parallel Sequencing in Clinical Diagnostics", Journal of Molecular Diagnostics, vol. 10, No. 6, 2008, 484-492.

Tewhey, R. et al., "Microdroplet-based PCR enrichment for large-scale targeted sequencing", Nature Biotechnology, vol. 27, No. 11, Nov. 2009, 1025-1031.

Tewhey, R. et al., "The importance of phase information for human genomics", Nature Reviews Genetics, vol. 12, No. 3, Mar. 1, 2011, 215-223.

Thavarajah, R. et al., "Chemical and physical basics of routine formaldehyde fixation", Journal of Oral and Maxillofacial Pathology, vol. 16, No. 3, 2012, 400-405.

The Bump Message Boards, The Bump (Panorama Test, attached), Jul. 1, 2013, 8 pages.

The International Hapmap Consort, "The International HapMap Project", Nature, vol. 426, Dec. 18, 2003, 789-796.

Thermofisher Scientific, "Ion AmpliSeq Cancer Hotspot Panel v2", Retrieved from the Internet: https://tools.thermofisher.com/content/sfs/brochures/Ion-AmpliSeq-Cancer-Hotspot-Panel-Flyer.pdf, 2015, 2 pages.

Thomas, M.R et al., "The Time of Appearance and Disappearance of Fetal DNA from the Maternal Circulation", Prenatal Diagnosis, 15, 1995, 641 -646.

Thompson, J. C. et al., "Detection of Therapeutically Targetable Driver and Resistance Mutations in Lung Cancer Patients by Next-Generation Sequencing of Cell-Free Circulating Tumor DNA", Clin Cancer Res, vol. 22, No. 23, Dec. 1, 2016, 5772-5782. Thornton, Brenda et al., "Real-time Per (qPCR) Primer Design

Thornton, Brenda et al., "Real-time Per (qPCR) Primer Design Using Free Online Software", Biochemistry and Molecular Biology Education, vol. 39, 2011, pp. 145-154.

Tie, et al., "Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer", Annals of Oncology vol. 26, No. 8, 2015, 1715-1722

Annals of Oncology, vol. 26, No. 8, 2015, 1715-1722.

Tiersch, T. R. et al., "Reference Standards for Flow Cytometry and Application in Comparative Studies of Nuclear DNA Content", Cytometry, vol. 10, Mar. 21, 1989, 706-710.

Cytometry, vol. 10, Mar. 21, 1989, 706-710.
Tong, et al., "Diagnostic Developments Involving Cell-free (Circulating) Nucleic Acids", Clinica Chimica Acta, vol. 363, No. (1-2), Aug. 26, 2005, 187-196.

Tong, Yu et al., "Noninvasive Prenatal Detection of Fetal Trisomy 18 by Epigenetic Allelic Ratio Analysis in Maternal Plasma: Theoretical and Empirical Considerations", Clinical Chemistry, 52(12), 2006, 2194-2202.

Tong, Yu K. et al., "Noninvasive Prenatal Detection of Trisomy 21 by Epigenetic-Genetic Chromosome-Dosage Approach", Clinical Chemistry 56(1) 2010 90-98

Chemistry, 56(1), 2010, 90-98. Toshikazu, et al., "Estimation of Haplotype Frequencies, Linkage-disequilibrium Measures, and Combination of Haplotype Copies in Each Pool by Use of Pooled DNA Data", American Journal of Human Genetics, vol. 72, Jan. 17, 2003, 384-398.

Toth, T. et al., "Prenatal Detection of Trisomy 13 From Amniotic Fluid by Quantitative Fluorescent Polymerase Chain Reaction", Prenatal Diagnosis, vol. 18, 1998, 669-674.

Tounta, G. et al., "Non-invasive prenatal diagnosis using cell-free fetal nucleic acids in maternal plasma: Progress overview beyond predictive and personalized diagnosis", EPMA Journal, vol. 2, Issue 2, 2011, 163-171.

Tounta, G. et al., "A Multiplex PCR for Non-invasive Fetal RHD Genotyping Using Cell-free Fetal DNA", in vivo, vol. 25, 2011, 411-418.

Treff, N. R. et al., "Single Cell Whole Genome Amplification Technique Significantly Impacts the Accuracy and Precision of Microarray Based 23 Chromosome Aneuploidy Screening", Poster Presentations Preimplantation Genetic Diagnosis, vol. 88, Supplement 1, Sep. 1, 2007, S231.

Troeger, C. et al., "Approximately Half of the Erythroblasts in Maternal Blood are of Fetal Origin", Molecular Human Reproduction, vol. 5, No. 12, Dec. 1, 1999, 1162-1165.

Troutt, et al., "Ligation-anchored PCR: A Simple Amplification Technique with Single-sided Specificity", Proceedings of the National Academy of Sciences, vol. 89, Oct. 1992, 9823-9825.

Troyanskaya, Olga G. et al., "A Bayesian Framework for Combining Heterogeneous Data Sources for Gene Function Prediction (in *Saccharomyces cerevisiae*)", PNAS, 100(14), 2003, 8348-8353.

Tsang, Jason C. et al., "Circulating Nucleic Acids in Plasma/ Serum", Pathology, vol. 39, No. 2, Apr. 1, 2007, 197-207.

Tsangaris, G. T. et al., "Proteomic analysis of amniotic fluid in pregnancies with Down syndrome", Proteomics, vol. 6, 2006, 4410-4419.

Tseng, Jeng-Sen et al., "Dynamic Plasma EGFR Mutation Status as a Predictor of EGFR-TKI Efficacy in Patients with fGFR-Mutant Lung Adenocarcinoma", Thorac Oncol., vol. 10, 2015, 603-610.

Tsui, N. B. et al., "Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling", J. Med. Genet, vol. 41, 2004, 461-467. Tsui, Nancy B.Y et al., "Non-Invasive Prenatal Detection of Fetal Trisomy 18 by RNA-SNP Allelic Ratio Analysis Using Maternal Plasma SERPINB2 mRNA: A Feasibility Study", Prenatal Diagnosis, 29, 2009, 1031-1037.

Tu, J. et al., "Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis", BMC Genomics, vol. 13, No. 43, Jan. 25, 2012, 1-9.

Tufan, N.L. et al., "Analysis of Cell-Free Fetal DNA from Maternal Plasma and Serum Using a Conventional Multiplex PCR: Factors Influencing Success", The Turkish Journal of Medical Sciences, vol. 35, 2005, 85-92.

Tungwiwat, et al., "Non-invasive fetal sex determination using a conventional nested PCR analysis of fetal DNA in maternal plasma", Clinica Chimica Acta, vol. 334, No. 1-2, 2003, 173-177.

Turner, E. et al., "Massively Parallel Exon Capture and Library-Free Resequencing Across 16 Genomes", Nature Methods, 6 (5), 2009, 315-316.

Tuzcu, et al., "Intravascular Ultrasound Evidence of Angiographically Silent Progression in Coronary Atherosclerosis Predicts Long-term Morbidity and Mortality After Cardiac Transplantation", The American Journal of Cardiology, vol. 45, No. 9, May 3, 2005, 1538-1542.

Tynan, J. A. et al., "Restriction Enzyme-Mediated Enhanced Detection of Circulating Cell-Free Fetal DNA in Maternal Plasma", The Journal of Molecular Diagnostics, vol. 13, No. 4, Jul. 2011, 382-389

Tzimagiorgis, G. et al., "Recovering circulating extracellular or cell-free RNA from bodily fluids", Cancer Epidemiology, vol. 35, 2011, 580-589.

Umetani, N. et al., "Increased Integrity of Free Circulating DNA in Sera of Patients with Colorectal or Periampullary Cancer: Direct Quantitative PCR for ALU Repeats", Clinical Chemistry, vol. 52, No. 6, 2006, 1062-1069.

Urbaniak, S. J. et al., "RhD haemolytic disease of the fetus and the newborn", Blood Reviews, vol. 14, 2000, 44-61.

Urbanova, M. et al., "Circulating Nucleic Acids as a New Diagnostic Tool", Cellular & Molecular Biology Letters, vol. 15, 2010, 242-259.

Vallone, P. M. et al., "A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome", Int J Legal Medicine, vol. 118, Feb. 4, 2004, 147-157.

Vallone, Peter , "AutoDimer: a Screening Tool for Primer-Dimer and Hairpin Structures", Bio Techniques, 37, 2004, 226-231. Vallone, Peter M. et al., "Demonstration of Rapid Multiplex PCR

Vallone, Peter M. et al., "Demonstration of Rapid Multiplex PCR Amplification Involving 16 Genetic Loci", Forensic Science International Genetics, vol. 3, 2008, pp. 42-45.

Van Den Oever, J. M. et al., "Single Molecule Sequencing of Free DNA from Maternal Plasma for Noninvasive Trisomy 21 Detection", Clinical Chemistry, vol. 58, No. 4, 2012, 699-706.

Van Uitert, I. et al., "The influence of different membrane components on the electrical stability of bilayer lipid membranes", Biochimica et Biophysica Acta, vol. 1798, 2010, 21-31.

## US 11,530,454 B2

Page 28

### (56) References Cited

#### OTHER PUBLICATIONS

Vanneste, Marion et al., "Functional Genomic Screening Independently Identifies CUL3 as a Mediator of Vemurafenib Resistance via Src-RAC1 Signaling Axis", Frontiers in Oncology, vol. 10, 2020, 16 pages.

Varley, Katherine Elena et al., "Nested Patch PCR Enables Highly Multiplexed Mutation Discovery in Candidate Genes", Genome Res., 18(11), 2008, 1844-1850.

Ventura-Aguiar, P. et al., "Donor-derived Cell-free DNA Shows High Sensitivity for the Diagnosis of Pancreas Graft Rejection in Simultaneous Pancreas-Kidney Transplantation", Transplantation, vol. 00, No. 00, 2022, 8 pages.

Verlaan, et al., "Allele-specific Chromatin Remodeling in the ZPBP22/GSDMB/ORMDL3 Locus Associated with the Risk of Asthma and Autoimmune Disease", The American Journal of Human Genetics, vol. 85, No. 3, Sep. 11, 2009, 377-393.

Verlaan, et al., "Targeted Screening of Cis-Regulatory Variation in Human Haplotypes", Genome Research, vol. 19, No. 1, Jan. 1, 2009, 118-127.

Verlinsky, Y. et al., "Over a Decade of Experience with Preimplantation Genetic Diagnosis", Fertility and Sterility, 82 (2), 2004, 302-303

Vlaminck, I. D. et al., "Circulating Cell-Free DNA Enables Noninvasive Diagnosis of Heart Transplant Rejection", Sci Transl Med., vol. 6, No. 241, Jun. 18, 2014, 26 pages.

Voelkerding, et al., "Next-generation Sequencing: From Basic Research to Diagnostics", Clinical Chemistry, vol. 55, No. 4, Apr. 1, 2009, 641-658.

Vogelstein, B. et al., "Digital PCR", Proc. Natl. Acad. Sci. USA, vol. 96, Aug. 1999, 9236-9241.

Volckmar, et al., "A field guide for cancer diagnostics using cell-free DNA: From principles to practice and clinical applications", Genes Chromosomes Cancer, 2018, 123-139.

Von Ahsen, Nicolas et al., "Oligonucleotide Melting Temperatures under PCR Conditions: Nearest-Neighbor Corrections for Mg2+, Deoxynucleotide Triphosphate, and Dimethyl Sulfoxide Concentrations with Comparison to Alternative Empirical Formulas", Clinical Chemistry, vol. 47, 2001, pp. 1956-1961.

Von Eggeling, F. et al., "Applications of Random PCR", Cellular and Molecular Biology, vol. 41, No. 5, 1995, 653-670.

and Molecular Biology, vol. 41, No. 5, 1995, 653-670. Wagner, F. F. et al., "RHD gene deletion occurred in the Rhesus box" Blood vol. 95 No. 12, 2000, 3662-3668.

box", Blood, vol. 95, No. 12, 2000, 3662-3668. Wagner, Jasenka et al., "Non-Invasive Prenatal Paternity Testing from Maternal Blood". Int. J. Legal Med., 123, 2009, 75-79.

from Maternal Blood", Int. J. Legal Med., 123, 2009, 75-79. Wang, et al., "Molecular inversion probes: a novel microarray technology and its application in cancer research", Cancer Genetics, 205, 2012, 341-355.

Wang, D. G. et al., "Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome", Science, vol. 280, May 15, 1998, 1077-1082. Wang, Eric et al., "Gestational Age and Maternal Weight Effects on

Wang, Eric et al., "Gestational Age and Maternal Weight Effects on Fetal Cell-Free DNA in Maternal Plasma", Prenatal Diagnosis, 33, 2013, 662-666.

Wang, Hui-Yun et al., "A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome", Genome Res., 15, 2005, 276-283.

Wang, J. et al., "Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm", Cell, vol. 150, Jul. 20, 2012, 402-412.

Wang, S. et al., "Potential Clinical Significance of a Plasma-Based KRAS Mutation Analysis in Patients with Advanced Non-Small Cell Lung Cancer", Clin Cancer Res, vol. 16, No. 4, Feb. 15, 2010, 1324-1330.

Wang, T.L. et al., "Digital karyotyping", PNAS, vol. 99, No. 25, Dec. 10, 2002, 16156-16161.

Wang, W.-P. et al., "Multiplex single nucleotide polymorphism genotyping by adapter ligation-mediated allele-specific amplification", Analytical Biochemistry, vol. 355, May 5, 2006, 240-248.

Wang, Yuker et al., "Allele quantification using molecular inversion probes (MIP)", Nucleic Acids Research, vol. 33, No. 21, Nov. 28, 2005, 14 pgs.

Wangkumhang, P. et al., "WASP: a Web-based Allele-Specific PCR assay designing tool for detecting SNPs and mutations", BMC Genomics, vol. 8, No. 275, Aug. 14, 2007, 9 pgs.

Wapner, R. et al., "Chromosomal Microarray Versus Karyotyping for Prenatal Diagnosis", The New England Journal of Medicine, 367 (23), 2012, 2175-2184.

Wapner, R. et al., "First-Trimester Screening for Trisomies 21 and 18", The New England Journal of Medicine, vol. 349, No. 15, Oct. 9, 2003, 1405-1413.

Wapner, R. J. et al., "Expanding the scope of noninvasive prenatal testing: detection of fetal microdeletion syndromes", American Journal of Obstetrics & Gynecology, vol. 212, Dec. 17, 2014, 1.e1-1.e9.

Wartell, Roger M. et al., "Thermal Denaturation of DNA Molecules: A Comparison of Theory with Experiment", Physics Reports, vol. 126, 1985, pp. 67-107.

Wasson, Jon et al., "Assessing Allele Frequencies of Single Nucleotide Polymorphisms in DNA Pools by Pyrosequencing Technology", BioTechniques, vol. 32, No. 5, May 1, 2002, 1144-1152.

Watkins, N. et al., "Thermodynamic contributions of single internal rA •dA, rC • dC, rG • dG and rU • dT mismatches in RNA/DNA duplexes", Nucleic Acids Research, 9 (5), 2010, 1894-1902.

Watt, Heather L., "Sex Diagnosis of Preimplantation Porcine Embryos through PCR Amplification of the Sry Gene", Sex Diagnosis of Preimplantation Porcine Embryos Through PCR Amplification of the SRY Gene (1998) ("Watt (1998)"), 1998, 151 pages. Wei, C. et al., "Detection and Quantification by Homogeneous PCR of Cell-free Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 47, No. 2, 2001, 336-338.

Wei, Ting et al., "Novel Approaches to Mitigate Primer Interaction and Eliminate Inhibitors in Multiplex PCR, Demonstrated Using an Assay for Detection of three Strawberry Viruses", Journal of Virological Methods, vol. 151, 2008, pp. 132-139.

Weiss, C. A., "Chapter 8: Confidence Intervals for One Population Mean", Introductory Statistics, Sixth Edition, 2002, 340-381.

Wellnhofer, et al., "Angiographic Assessment of Cardiac Allograft Vasculopathy: Results of a Consensus Conference of the Task Force for Thoracic Organ Transplantation of the German Cardiac Society", Transplant International, vol. 23, No. 11, Aug. 19, 2010, 1094-1104.

Wells, D , "Microarray for Analysis and Diagnosis of Human Embryos", 12th International Congress on Prenatal Diagnosis and Therapy, Budapest, Hungary, 2004, 9-17.
Wells, Dagan, "Advances in Preimplantation Genetic Diagnosis",

Wells, Dagan, "Advances in Preimplantation Genetic Diagnosis", European Journal of Obstetrics and Gynecology and Reproductive Biology, 115S, 2004, S97-S101.

Wells, Dagan, "Detailed Chromosomal and Molecular Genetic

Wells, Dagan, "Detailed Chromosomal and Molecular Genetic Analysis of Single Cells by Whole Genome Amplification and Comparative Genomic Hybridisation", Nucleic Acids Research, 27, 4, 1999, 1214-1218.

Wen, Daxing et al., "Universal Multiples PCR: A Novel Method of Simultaneous Amplification of Multiple DNA Fragments", Plant Methods, 8(32), NULL, 2012, 1-9.

What To Expect Message Boards, What To Expect (Weird Harmony results), May 1, 2015, 7 pages.
Whitlam, J. B. et al., "Diagnostic application of kidney allograft-

Whitlam, J. B. et al., "Diagnostic application of kidney allograftderived absolute cell-free DNA levels during transplant dysfunction", Am J Transplant, vol. 19, 2019, 1037-1049.

Widlak, P. et al., "Cleavage Preferences of the Apoptotic Endonuclease DFF 40 (Caspase~activated DNase or Nuclease) on Naked DNA and Chromatin Substrates", The Journal of Biological Chemistry, vol. 275, No. 11, Mar. 17, 2000, 8228-8232.

Wiedmann, Ralph T. et al., "SNP Discovery in Swine by Reduced Representation and High Throughput Pyrosequencing", BMC Genetics, vol. 9, Article No. 81, Dec. 4, 2008, 1-7.

Wikipedia, "Buffy coat", Retrieved from "https://en.wikipedia.org.Jw/index.php?title=Buffy\_coat&oldid=900992886", Jun. 9, 2019, 2 pgs.

# US 11,530,454 B2

Page 29

#### (56) References Cited

#### OTHER PUBLICATIONS

Wikipedia, "Maximum a posteriori estimation", https://en.wikipedia.org/w/index.php?title=Maximum\_a\_posteriori\_estimation&oldid=26878808, [retrieved on Aug. 1, 2017], Oct. 30, 2005, 2 pages. Wikipedia, "Stimulant", 2016, 17 pages.

Wilkening, Stefan et al., "Determination of Allele Frequency in Pooled DNA: Comparison of Three PCR-based Methods", Bio Techniques, vol. 39, No. 6, May 30, 2005, 853-857.

Wilkinson, Sarah T. et al., "Decreased MHC Class II Expression in Diffuse Large B-Cell Lymphoma does not Correlate with CPG Methylation of Ciita Promoters III and IV", Leuk Lymphoma, vol. 50, 2009, pp. 1875-1878.

Wilton, et al., "Birth of a Healthy Infant After Preimplantation Confirmation of Euploidy by Comparative Genomic Hybridization", N. Engl. J. Med., 345(21), 2001, 1537-1541.

Wilton, L., "Preimplantation Genetic Diagnosis and Chromosome Analysis of Blastomeres Using Comparative Genomic Hybridization", Human Reproduction Update, 11 (1), 2005, 33-41.

Winsor, E. J. et al., "Maternal Cell Contamination in Uncultured Amniotic Fluid", Prenatal Diagnosis, vol. 16, 1996, 49-54.

Witherspoon, David J. et al., "Mobile Element Scanning (Me-scan) by Targeted High-throughput Sequencing", BMC Genomics, vol. 410, 2010, 15 pages.

Wittwer, C. T. et al., "Real-Time Multiplex PCR Assays", Methods, vol. 25, 2001, 430-448.

Wong, K. H. et al., "Multiplex Illumina Sequencing Using DNA Barcoding", Current Protocols in Molecular Biology, vol. 101, Jan. 2013, 7.11.1-7.11.11.

Wong, K. K. et al., "Allelic imbalance analysis by high-density single nucleotide polymorphic allele (SNP) array with whole genome amplified DNA", Nucleic Acids Research, vol. 32, No. 9, May 17, 2004, 8 pages.

Wright, C. et al., "The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis", Human Reproduction Update, vol. 15, No. 1, 2009, 139-151.

Wright, C. F. et al., "Cell-free fetal DNA and RNA in maternal blood implications for safer antenatal testing", BMJ, vol. 39, Jul. 18, 2009, 161, 165

Wright, Caroline et al., "Cell-free Fetal Nucleic Acids for Noninvasive Prenatal Diagnosis", PHG Foundation, Jan. 1, 2009, 1-64. Wu, T.L. et al., "Cell-free DNA: measurement in various carcinomas and establishment of normal reference range", Clinica Chimica Acta, vol. 321, 2002, 77-87.

Wu, Y. Y. et al., "Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications", Clinica Chimica Acta, vol. 363, 2006, 165-176.

applications", Clinica Chimica Acta, vol. 363, 2006, 165-176. Xia, et al., "Simultaneous Quantitative Assessment of Circulating Cell-free Mitochondrial and Nuclear DNA by Multiplex Real-time PCR", Genetics and Molecular Biology, vol. 32, No. 1, Mar. 1, 2009, 20-24.

Xia, Tianbing et al., "Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs", Biochemistry, 37, 1998, 14719-14735. Xian, et al., "Advances on Circulating Fetal DNA in Maternal Plasma", Chinese Medical Journal, vol. 120, No. 14, Jul. 2, 2007, 1256-1259

Xie, et al., "CNV-SEQ, a New Method to Detect Copy Number Variation Using Highthroughput Sequencing", BMC Bioinformatics, vol. 10:80, Mar. 6, 2009, 1-9.

Xu, N. et al., "A Mutation in the Fibroblast Growth Factor Receptor 1 Gene Causes Fully Penetrant Normosmic Isolated Hypogonadotropic Hypogonadism", The Journal of Clinical Endocrinology & Metabolism, vol. 92, No. 3, 2007, 1155-1158.

Xu, S. et al., "Circulating tumor DNA identified by targeted sequencing in advanced-stage non-small cell lung cancer patients", Cancer Letters, vol. 370, 2016, 324-331.

Xu, W. et al., "A Novel Universal Primer-Multiplex-PCR Method with Sequencing Gel Electrophoresis Analysis", PLOS One, vol. 7, No. 1, Jan. 17, 2012, 10 pgs.

Xue, et al., "Optimizing the Yield and Utility of Circulating Cellfree DNA From Plasma and Serum", Clinica Chimica Acta, vol. 404, No. 2, Jun. 27, 2009, 100-104.

Yamada, T. et al., "Detection of K-ras Gene Mutations in Plasma DNA of Patients with Pancreatic Adenocarcinoma: Correlation with Clinicopathological Features", Clinical Cancer Research, vol. 4, Jun. 1998, 1527-1532.

Yamada, T. et al., "PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome", Nucleic Acids Research, vol. 34, 2006, W665-W669.

Yang, Lin et al., "64-MDCT Coronary Angiography of Patients With Atrial Fibrillation: Influence of Heart Rate on Image Quality and Efficacy in Evalution of Coronary Artery Disease", AJR, vol. 193, No. 3, Sep. 1, 2009, 795-801.

Yaron, Y., "The implications of non-invasive prenatal testing fail-

Yaron, Y., "The implications of non-invasive prenatal testing failures: a review of an under-discussed phenomenon", Prenatal Diagnosis, vol. 36, 2016, 391-396.

Ye, et al., "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction", BMC Bioinformatics, 13:134, 2012, 11 pages.

Yeh, Iwei et al., "Knowledge Acquisition, Consistency Checking and Concurrency Control for Gene Ontology (GO)", Bioinformatics, 19, 2, 2003, 241-248.

Yijen, et al., "Noninvasive Evaluation of Cardiac Allograft Rejection by Cellular And Functional Cardiac Magnetic Resonance", JACC Cardiovacular Imaging, vol. 2, No. 6, Jun. 1, 2009, 731-741. Yilmaz, A. et al., "Comparative Evaluation of Left and Right Ventricular Endomyocardial Biopsy", Circulation, vol. 122, No. 9, Aug. 31, 2010, 900-909.

You, Frank M. et al., "BatchPrimerS: A high throughput web application for PCR and sequencing primer design", BMC Bioinformatics, Biomed Central, London, GB, vol. 9, No. 1, May 29, 2008 (May 29, 2008), p. 253.

Yuan, X. et al., "Probability Theory-based SNP Association Study Method for Identifying Susceptibility Loci and Genetic Disease Models in Human Case-Control Data", IEEE Trans Nanobioscience, vol. 9, No. 4, Dec. 2010, 232-241.

Yuanxin, Yan et al., "T-linker-specific Ligation PCR (T-linker Per): An Advanced PCR Technique for Chromosome Walking or for Isolation of Tagged DNA Ends", NucleicAcids Research, vol. 31, No. 12, e68, 2003, 7 pages.

Yung, T. K. et al., "Single-Molecule Detection of Epidermal Growth

Yung, T. K. et al., "Single-Molecule Detection of Epidermal Growth Factor Receptor Mutations in Plasma by Microfluidics Digital PCR in Non-Small Cell Lung Cancer Patients", Clinical Cancer Research, vol. 15, Mar. 10, 2009, 2076-2084.

Zachariah, R. et al., "Circulating cell-free DNA as a potential

Zachariah, R. et al., "Circulating cell-free DNA as a potential biomarker for minimal and mild endometriosis", Reproductive BioMedicine Online, vol. 18, No. 3, Jan. 27, 2009, 4007-411.

Zhang, et al., "Diagnosis of Acute Rejection by Analysis of Urinary DNA of Donor Origin in Renal Transplant Recipients", Transplantation Proceedings, vol. 33, No. 1-2, Feb. 2001, 380-381. Zhang, et al., "Use of PCR And PCR-SSP for Detection of Urinary

Zhang, et al., "Use of PCR And PCR-SSP for Detection of Urinary Donor-Origin Dna in Renal Transplant Recipients With Acute Rejection", Chinese Medical Journal, vol. 116, No. 2, Feb. 2003, 191-194.

Zhang, J. et al., "Presence of Donor-and Recipient-derived DNA in Cell-free Urine Samples of Renal Transplantation Recipients: Urinary DNA Chimerism", Clinical Chemistry, vol. 45, No. 10, 1999, 1741-1746.

Zhang, Kun et al., "Digital RNA Alleotyping Reveals Tissue-specific and Allele-specific Gene Expression in Human", Nature Methods, vol. 6, No. 8, Jul. 20, 2009, 613-618.

Zhang, L. et al., "Whole genome amplification from a single cell Implications for genetic analysis", Proc. Nat'l. Acad. Sci. USA, vol. 89, Jul. 1992, 5847-5851.

Zhang, Rui et al., "Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing", Nature Methods, 11(1), 2014, 51-56

Zhao, et al., "Urinary Thromboxane B2 in Cardiac Transplant Patients as a Screening Method of Rejection", Prostaglandins, vol. 54, No. 6, Dec. 1, 1997, 881-889.

Case: 24-1324 Document: 42-1 Page: 344 Filed: 03/18/2024

# US 11,530,454 B2

Page 30

#### **References Cited** (56)

#### OTHER PUBLICATIONS

Zhao, Xiaojun et al., "An Integrated View of Copy Number and Allelic Alterations in the Cancer Genome Using Single Nucleotide

Allelic Alterations in the Cancer Genome Using Single Nucleotide Polymorphism Arrays", Cancer Research,64, 2004, 3060-3071. Zheng, S. et al., "Whole Genome Amplification Increases the Efficiency and Validity of Buccal Cell Genotyping in Pediatric Populations!", Cancer Epidemiology, Biomarkers & Prevention, vol. 10, Jun. 2001, 697-700. Zheng, Z et al., "Anchored Multiplex PCR for Targeted Next-

generation Sequencing", Nature Medicine, vol. 20, No. 12, Dec. 2014, 1479-1486.

Zhong, X Y. et al., "Detection of Fetal Rhesus D and Sex Using Fetal DNA from Maternal Plasma by Multiplex Polymerase Chain Reaction", British Journal of Obstetrics and Gynaecology, vol. 107, Jun. 2000, 766-769

Zhong, X. et al., "Risk free simultaneous prenatal identification of fetal Rhesus D status and sex by multiplex real-time PCR using cell free fetal DNA in maternal plasma", Swiss Medical Weekly, vol. 131, Mar. 2001, 70-74.

Zhong, Xiao Y. et al., "Cell-free DNA In Urine: A Marker for Kidney Graft Rejection, but Not for Prenatal Diagnosis?", Annals of the New York Academy of Sciences, vol. 945, Sep. 1, 2001,

Zhou, et al., "Pyrosequencing, a High-throughput Method for Detecting Single Nucleotide Polymorphisms in the Dihydrofolate Reductase and Dihydropteroate Synthetase Genes of Plasmodiym Falciparum", Journal of Clinical Microbiology, vol. 44, No. 11, Nov. 1, 2006, 3900-3910.

Zhou, W. et al., "Counting Alleles Reveals a Connection Between Chromosome 18q Loss and Vascular Invasion", Nature Biotechnology, 19, 2001, 78-81.

Zhou, W. et al., "Counting alleles to predict recurrence of earlystage colorectal cancers", The Lancet, vol. 359, Jan. 19, 2002,

Zimmer, et al., "Transplant Coronary Artery Disease", JACC: Cardiovascular Interventions, vol. 3, No. 4, Apr. 1, 2010, 367-377. Zimmermann, et al., "Noninvasive Prenatal Aneuploidy Testing of Chromosomes 13, 18, 21 X, and Y, Using targeted Sequencing of Polymorphic Loci", Prenatal Diagnosis, 32, 2012, 1-9.

Zimmermann, B., "Declaration Under 37 CFR 1.32", filed in U.S. Appl. No. 14/171,587, filed Feb. 3, 2014, 4 pgs.

Zimmermann, B. et al., "Digital PCR: a powerful new tool for noninvasive prenatal diagnosis?", Prenatal Diagnosis, vol. 28, Nov. 10, 2008, 1087-1093

Zimmermann, B., "Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci, Supplemental Information", Prenatal Diagnosis, vol. 32, 2012, 7 pages.

Zimmermann, B. et al., "Novel Real-Time Quantitative PCR Test for Trisomy 21", Clinical Chemistry, vol. 48, No. 2, 2002, 362-363. Zimmermann, B. et al., "Optimized Real-Time Quantitative PCR Measurement of Male Fetal DNA in Maternal Plasma", Clinical Chemistry, vol. 51, No. 9, 2005, 1598-1604. Zimmermann, B. et al., "Real-Time Quantitative Polymerase Chain

Reaction Measurement of Male Fetal DNA in Maternal Plasma", Methods in Molecular Medicine, vol. 132, 2007, 43-49.

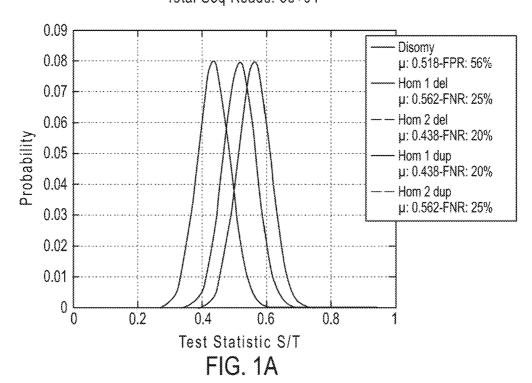
Zimmermann, B. et al., "Use of Real-Time Polymerase Chain Reaction for the Detection of Fetal Aneuploidies", Methods in Molecular Biology, vol. 336, Feb. 2006, 83-100.

Zlotogora, J., "Penetrance and expressivity in the molecular age", Genetics in Medicine, vol. 5, No. 5, 2003, 347-352.

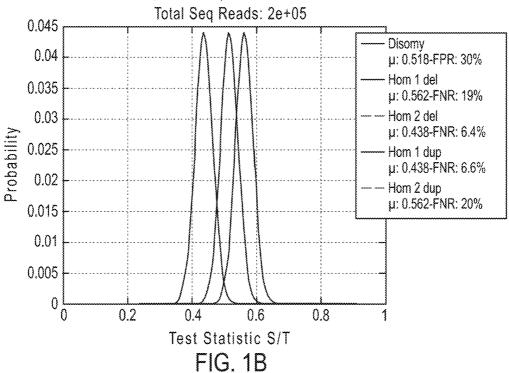
\* cited by examiner

# U.S. Patent Dec. 20, 2022 Sheet 1 of 105 US 11,530,454 B2

Tumor Fraction: 1%, Number of SNPs: 100 Total Seq Reads: 5e+04

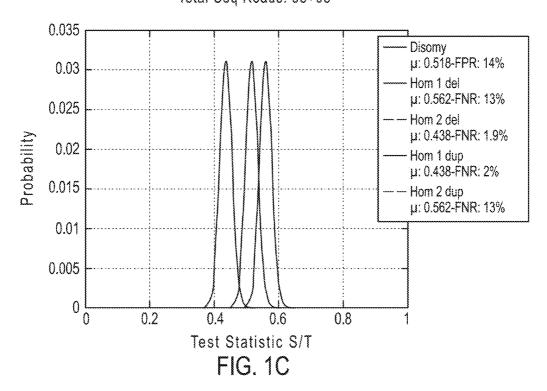


Tumor Fraction: 1%, Number of SNPs: 333

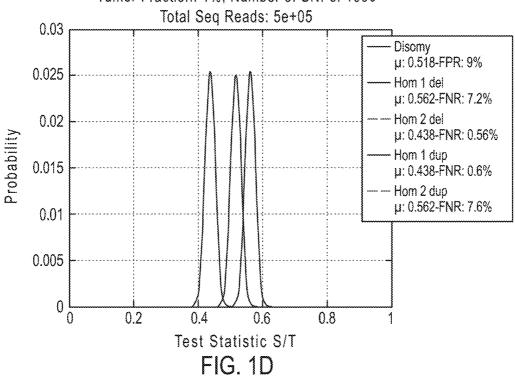


# U.S. Patent Dec. 20, 2022 Sheet 2 of 105 US 11,530,454 B2

Tumor Fraction: 1%, Number of SNPs: 667 Total Seq Reads: 3e+05



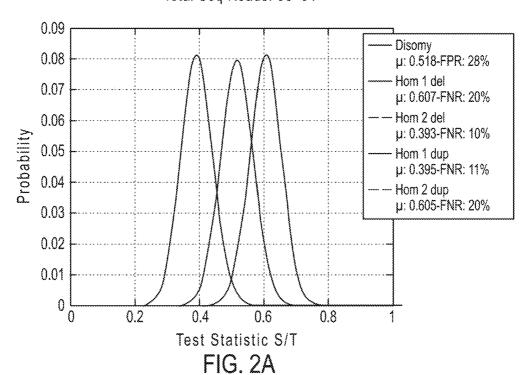
Tumor Fraction: 1%, Number of SNPs: 1000



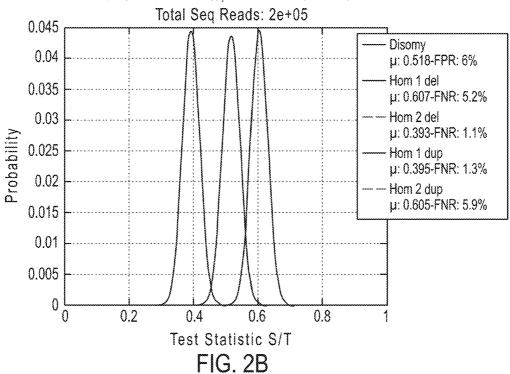
Document: 42-1 Page: 347 Case: 24-1324 Filed: 03/18/2024

U.S. Patent Dec. 20, 2022 **Sheet 3 of 105** US 11,530,454 B2

> Tumor Fraction: 2%, Number of SNPs: 100 Total Seq Reads: 5e+04



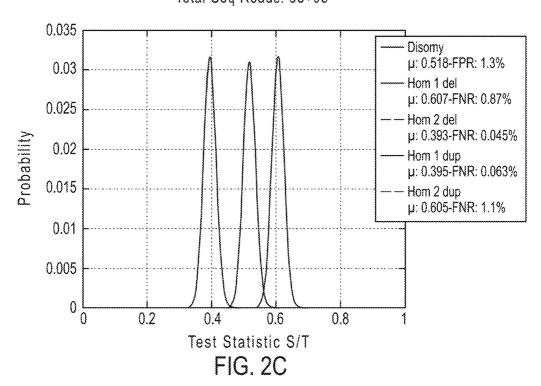
Tumor Fraction: 2%, Number of SNPs: 333 Total Seg Reads: 2e+05



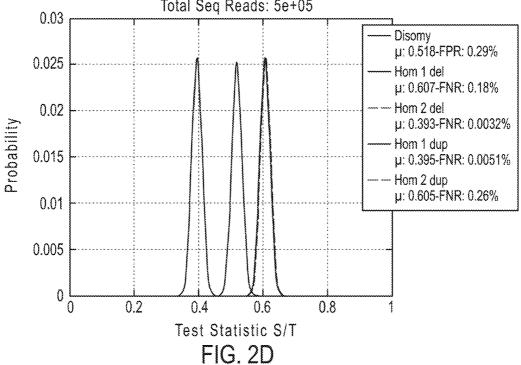
Document: 42-1 Page: 348 Case: 24-1324 Filed: 03/18/2024

#### U.S. Patent Dec. 20, 2022 **Sheet 4 of 105** US 11,530,454 B2

Tumor Fraction: 2%, Number of SNPs: 667 Total Seq Reads: 3e+05

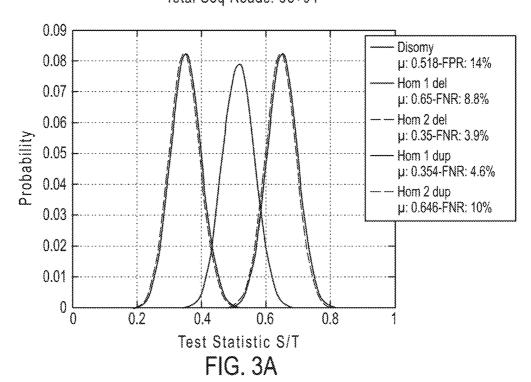


Tumor Fraction: 2%, Number of SNPs: 1000 Total Seq Reads: 5e+05



# U.S. Patent Dec. 20, 2022 Sheet 5 of 105 US 11,530,454 B2

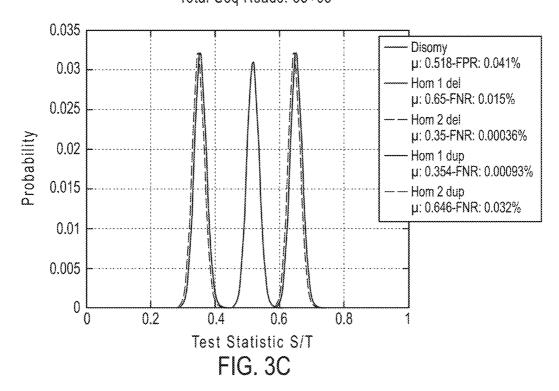
Tumor Fraction: 3%, Number of SNPs: 100 Total Seq Reads: 5e+04



Tumor Fraction: 3%, Number of SNPs: 333 Total Seg Reads: 2e+05 0.05 Disomy 0.045 μ: 0.518-FPR: 0.87% Hom 1 del 0.04 μ: 0.65-FNR: 0.63% 0.035 Hom 2 del µ: 0.35-FNR: 0.077% Probability 0.03 Hom 1 dup 0.025 u: 0,354-FNR: 0.13% Hom 2 dup 0.02 u: 0.646-FNR: 0.93% 0.015 0.01 0.005 0 0.2 0.6 0.8 Test Statistic S/T FIG. 3B

# U.S. Patent Dec. 20, 2022 Sheet 6 of 105 US 11,530,454 B2

Tumor Fraction: 3%, Number of SNPs: 667 Total Seq Reads: 3e+05



Tumor Fraction: 3%, Number of SNPs: 1000 Total Seq Reads: 5e+05 0.03 Disomy µ: 0.518-FPR: 0.0015% 0.025 Hom 1 del μ: 0.65-FNR: 0.00065% Hom 2 del 0.02 Probability μ: 0.35-FNR: 1.7e-06% Hom 1 dup 0.015 μ: 0.354-FNR: 6.8e-06% Hom 2 dup μ: 0.646-FNR: 0.002% 0.01 0.005 0 0.2 0.6 0.8 Test Statistic S/T FIG. 3D

Document: 42-1 Page: 351 Case: 24-1324 Filed: 03/18/2024

U.S. Patent Dec. 20, 2022 **Sheet 7 of 105** US 11,530,454 B2

> Tumor Fraction: 4%, Number of SNPs: 100 Total Seq Reads: 5e+04

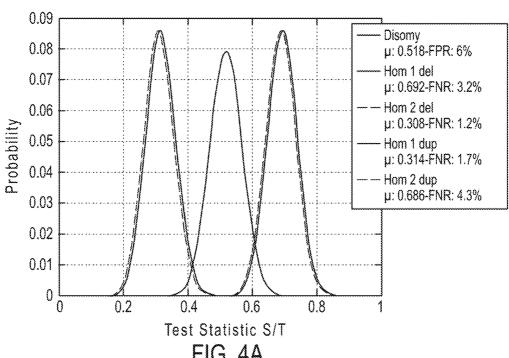
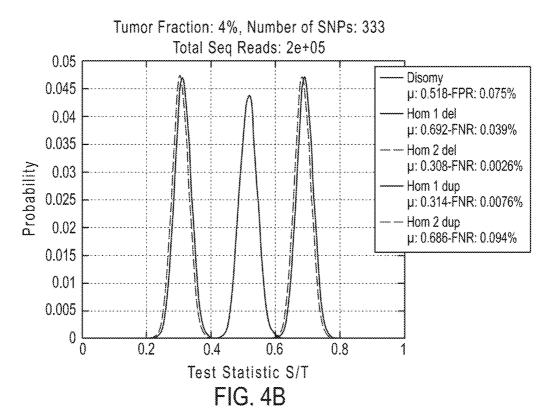
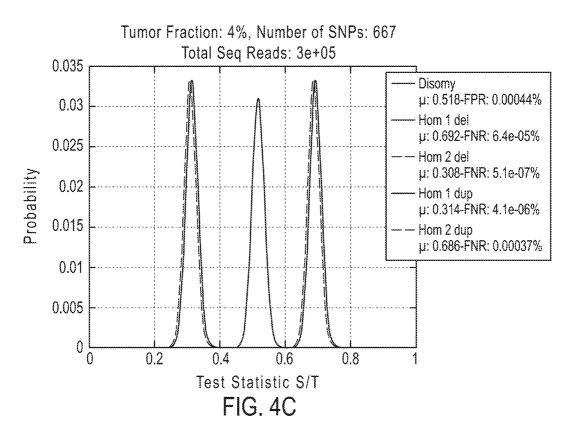
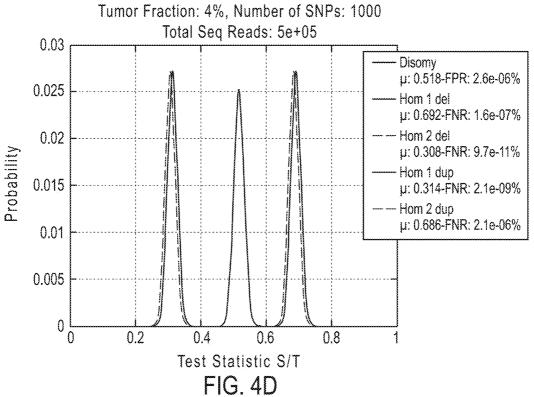


FIG. 4A

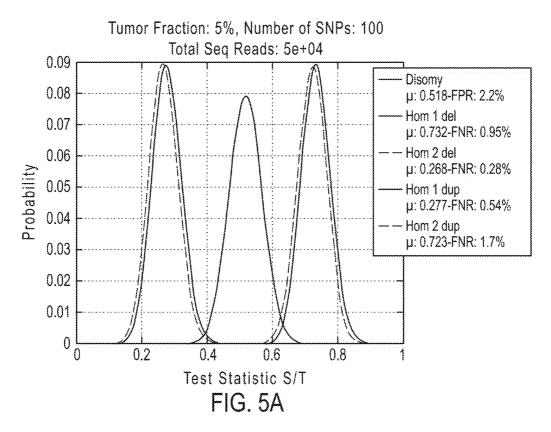


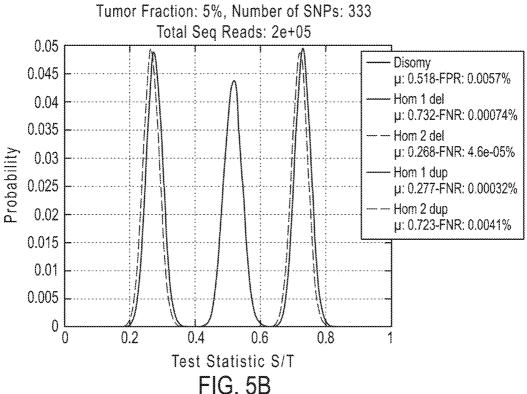
U.S. Patent Dec. 20, 2022 Sheet 8 of 105 US 11,530,454 B2



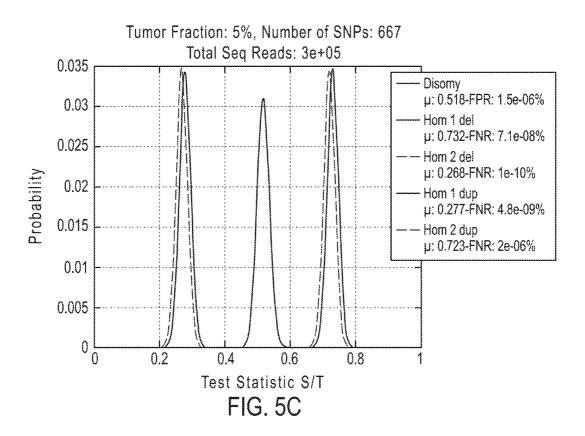


U.S. Patent Dec. 20, 2022 Sheet 9 of 105 US 11,530,454 B2



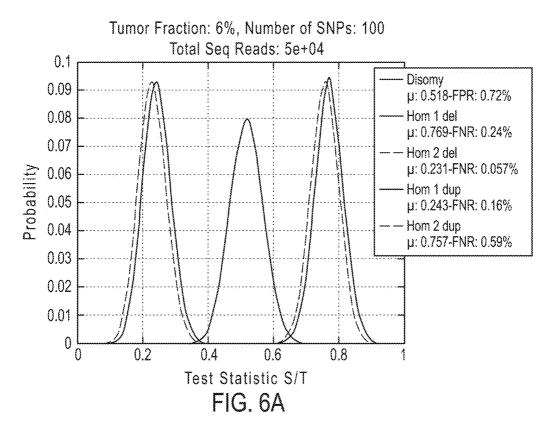


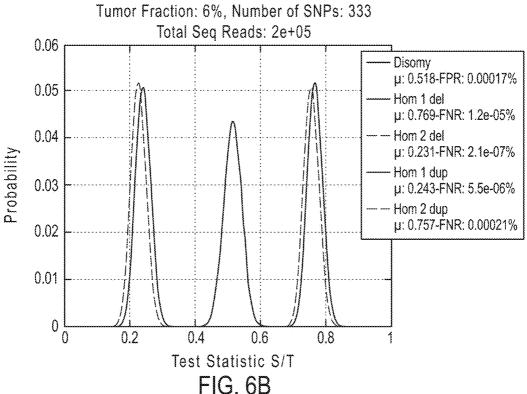
# U.S. Patent Dec. 20, 2022 Sheet 10 of 105 US 11,530,454 B2



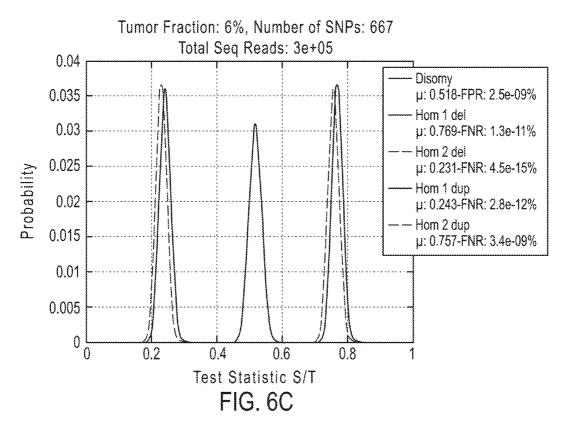
Tumor Fraction: 5%, Number of SNPs: 1000 Total Seq Reads: 5e+05 0.03 Disomy μ: 0.518-FPR: 8.5e-10% 0.025 Hom 1 del μ: 0.732-FNR: 4e-12% - Hom 2 del 0.02 Probability μ: 0.268-FNR: 3.8e-16% - Hom 1 dup 0.015 μ: 0.277-FNR: 1.2e-13% - Hom 2 dup μ: 0.723-FNR: 5.7e-10% 0.01 0.005 0 0.2 0.4 0.6 8.0 Test Statistic S/T FIG. 5D

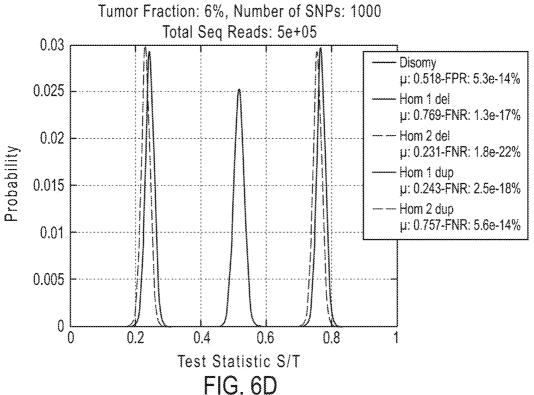
# U.S. Patent Dec. 20, 2022 Sheet 11 of 105 US 11,530,454 B2



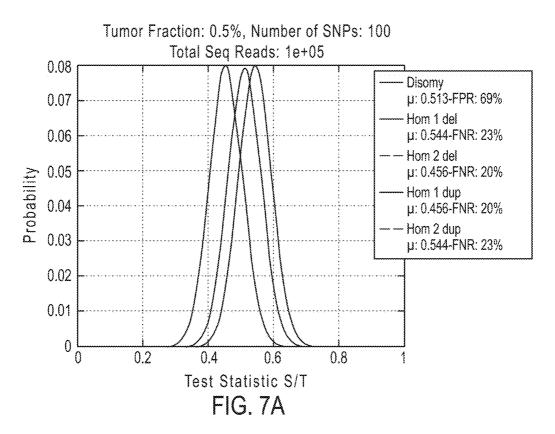


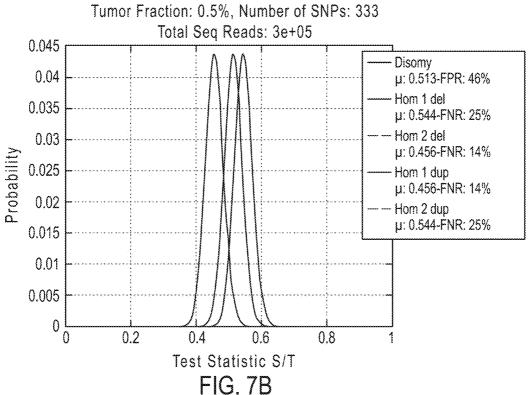
U.S. Patent Dec. 20, 2022 Sheet 12 of 105 US 11,530,454 B2



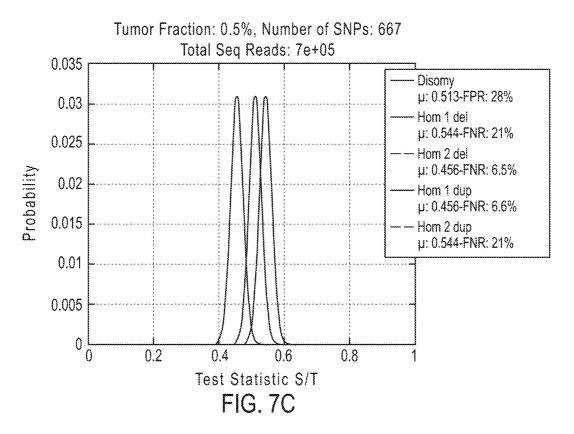


U.S. Patent Dec. 20, 2022 Sheet 13 of 105 US 11,530,454 B2



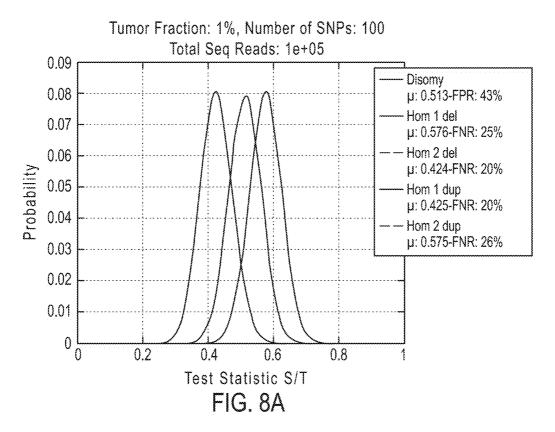


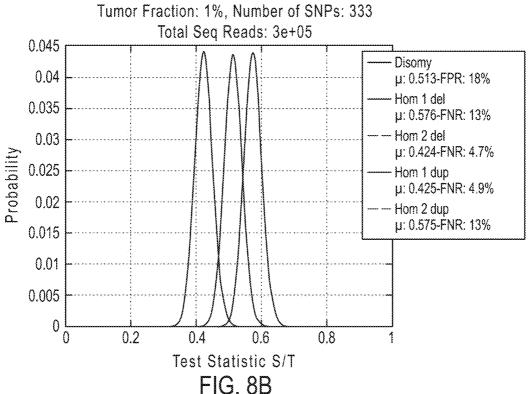
# U.S. Patent Dec. 20, 2022 Sheet 14 of 105 US 11,530,454 B2



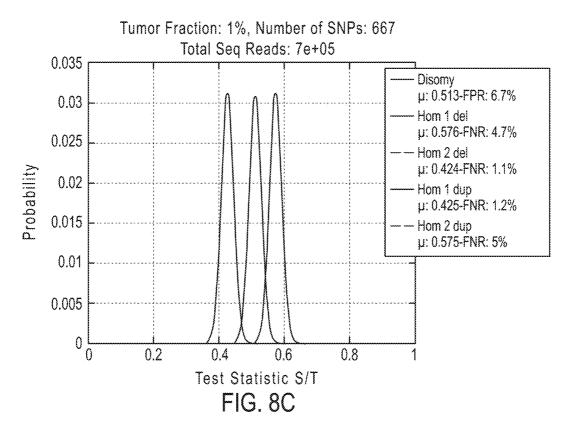
Tumor Fraction: 0.5%, Number of SNPs: 1000 Total Seq Reads: 1e+05 0.03 Disomy μ: 0.513-FPR: 20% 0.025 Hom 1 del μ: 0.544-FNR: 16% Hom 2 del 0.02 Probability μ: 0.456-FNR: 3.5% Hom 1 dup 0.015 u: 0.456-FNR: 3.5% - Hom 2 dup μ: 0.544-FNR: 16% 0.01 0.005 0 0 0.2 0.4 0.6 8.0 Test Statistic S/T FIG. 7D

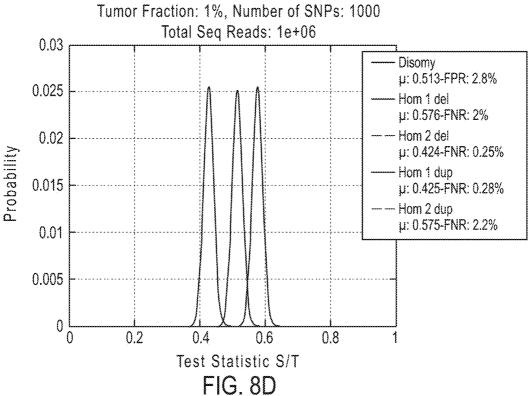
# U.S. Patent Dec. 20, 2022 Sheet 15 of 105 US 11,530,454 B2



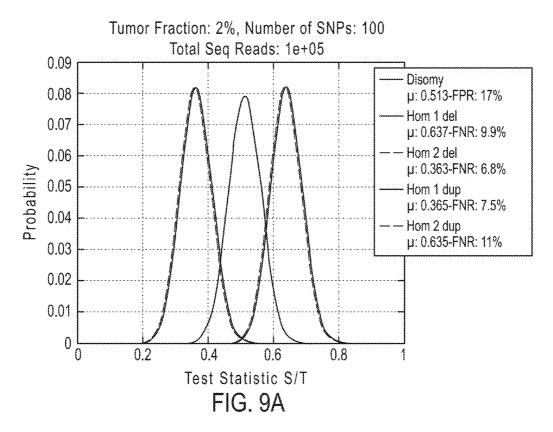


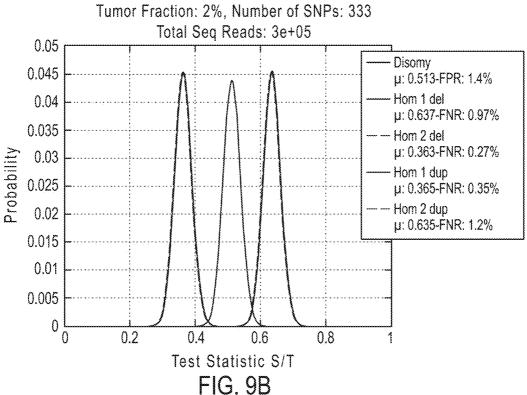
# U.S. Patent Dec. 20, 2022 Sheet 16 of 105 US 11,530,454 B2



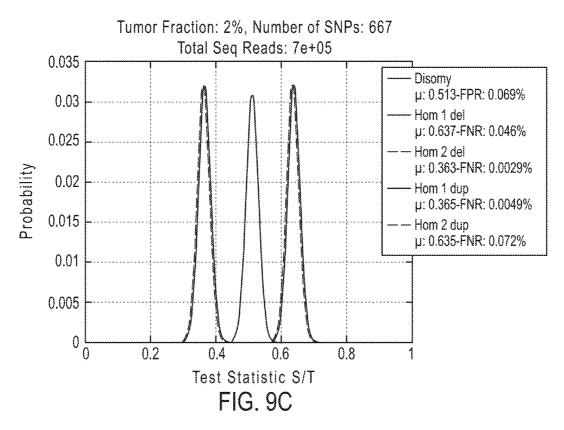


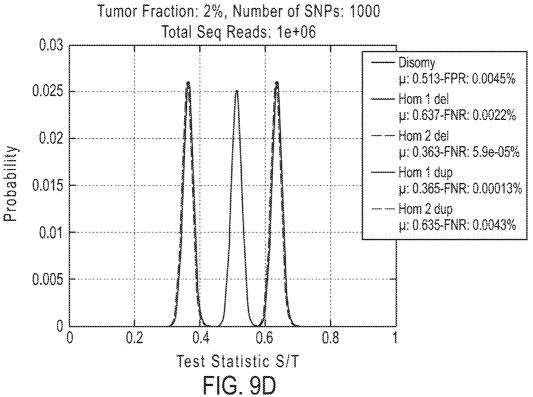
U.S. Patent Dec. 20, 2022 Sheet 17 of 105 US 11,530,454 B2



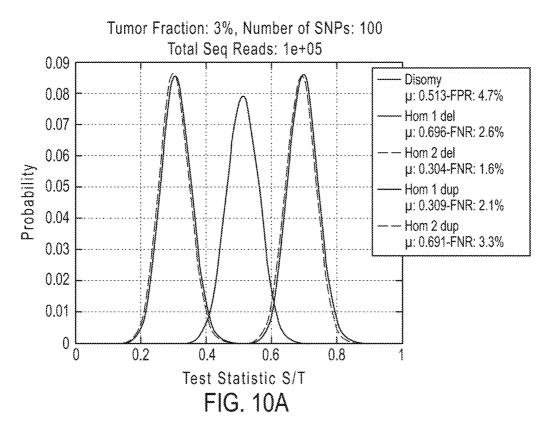


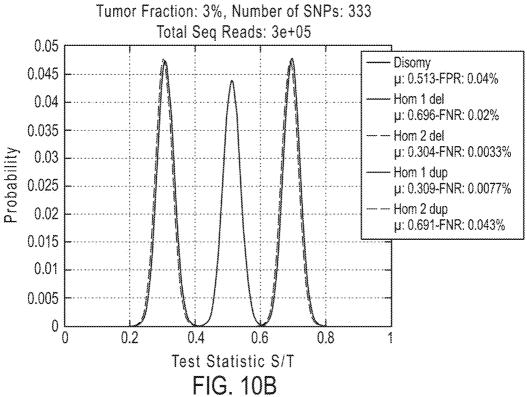
U.S. Patent Dec. 20, 2022 Sheet 18 of 105 US 11,530,454 B2



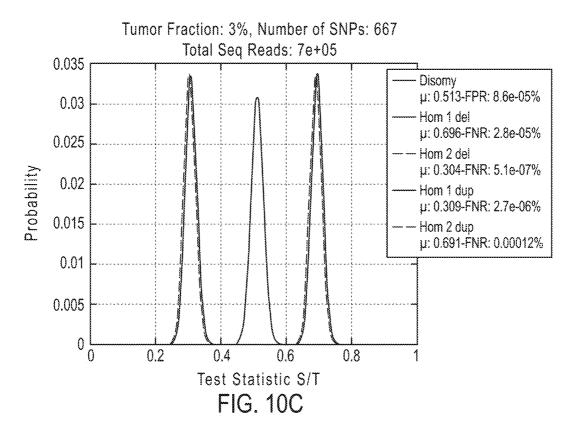


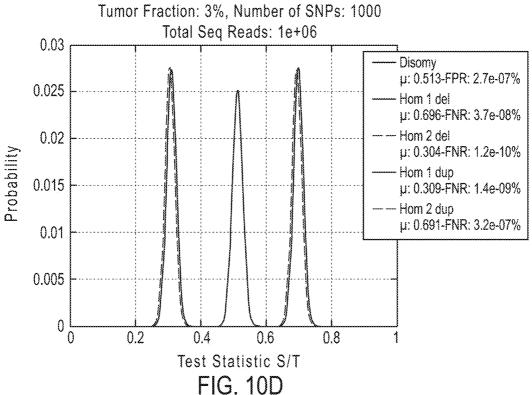
U.S. Patent Dec. 20, 2022 Sheet 19 of 105 US 11,530,454 B2



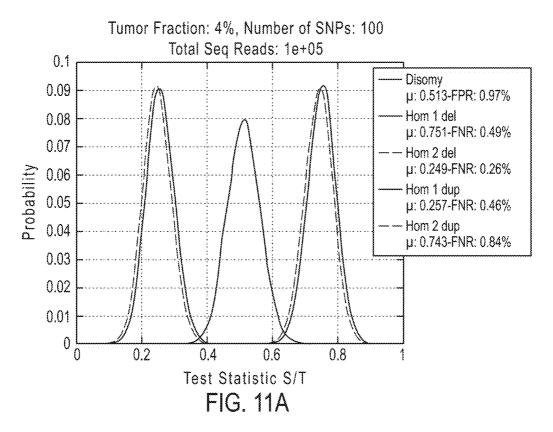


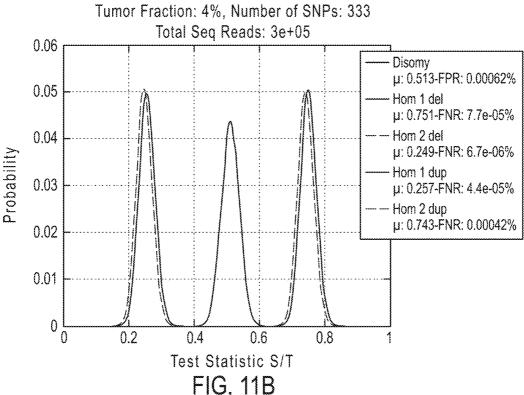
U.S. Patent Dec. 20, 2022 Sheet 20 of 105 US 11,530,454 B2



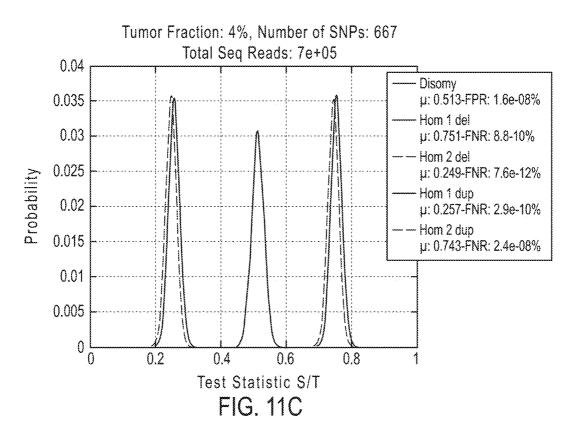


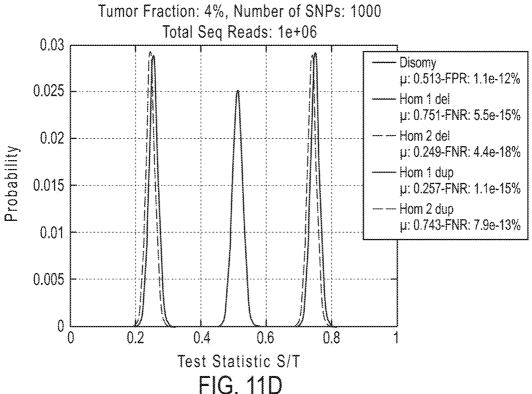
U.S. Patent Dec. 20, 2022 Sheet 21 of 105 US 11,530,454 B2



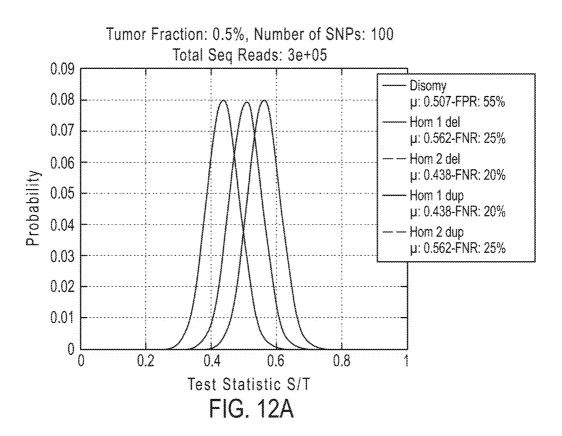


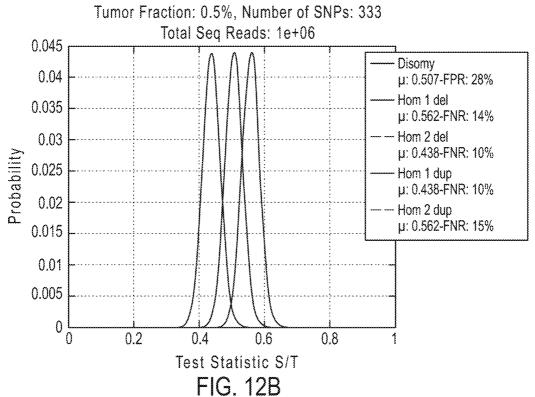
U.S. Patent Dec. 20, 2022 Sheet 22 of 105 US 11,530,454 B2





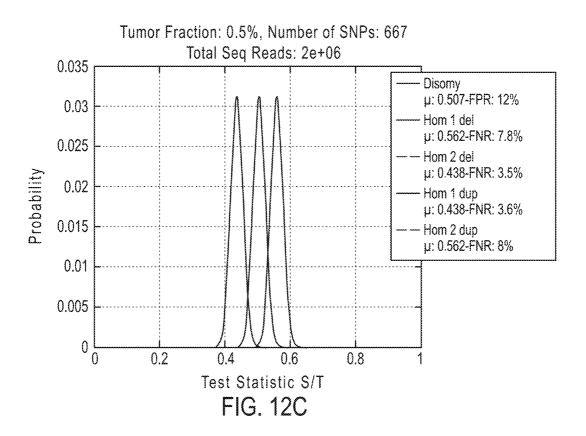
U.S. Patent Dec. 20, 2022 Sheet 23 of 105 US 11,530,454 B2





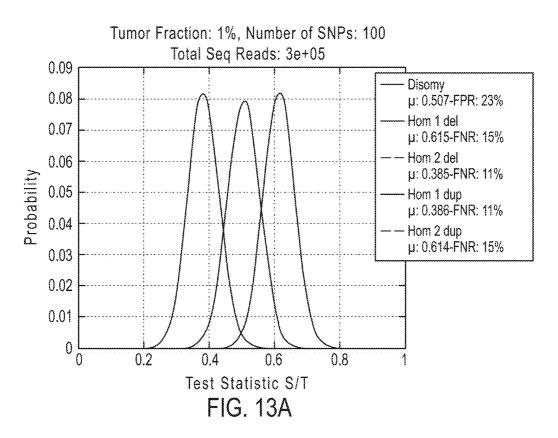
Document: 42-1 Case: 24-1324 Page: 368 Filed: 03/18/2024

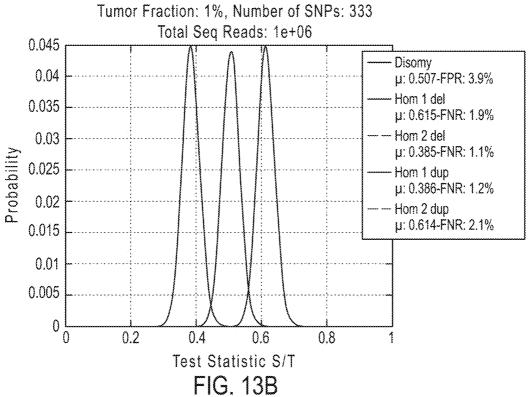
## U.S. Patent Dec. 20, 2022 **Sheet 24 of 105** US 11,530,454 B2



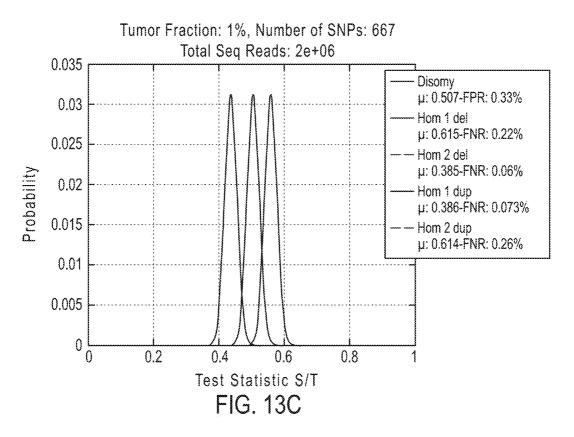
Tumor Fraction: 0.5%, Number of SNPs: 1000 Total Seq Reads: 3e+06 0.03 Disomy

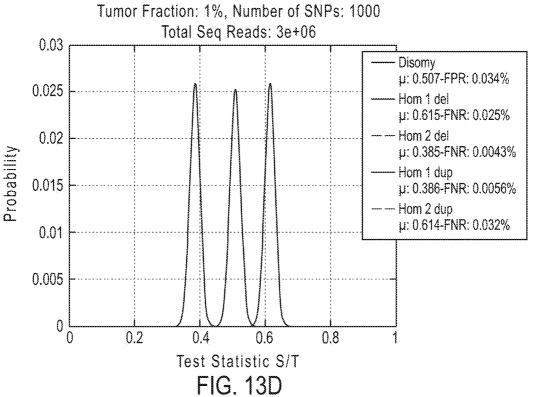
U.S. Patent Dec. 20, 2022 Sheet 25 of 105 US 11,530,454 B2





U.S. Patent Dec. 20, 2022 Sheet 26 of 105 US 11,530,454 B2



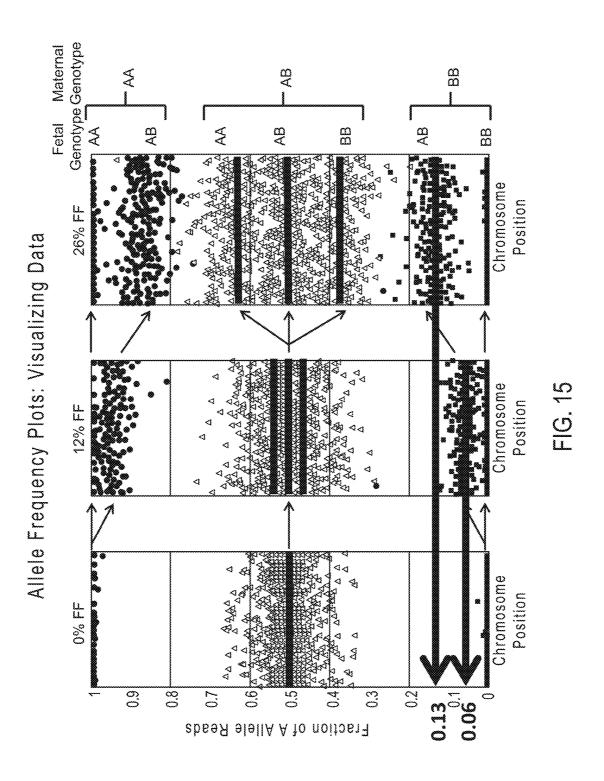


U.S. Patent Dec. 20, 2022 Sheet 27 of 105 US 11,530,454 B2

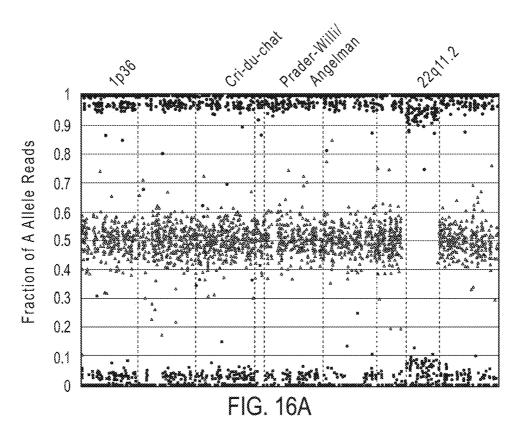
Syndrome	Sensitivity	Specificity
22all 2 Deletion / DiCoorge	97.8% (45/46)	99.2% (392/395)
22q11.2 Deletion/ DiGeorge	[CI: 88.5 – 99.95%]	[CI: 97.8 - 99.8%]
Angelman	100% (21/21)	100% (418/418)
	[CI: 83.9 – 100%]	[CI: 99.1 – 100%]
Cri du chat	100% (24/24)	99.8% (415/416)
	[CI: 85.7 – 100%]	[CI: 98.7 – 99.99%]
Monosomy 1p36	100% (1/1)	100% (438/438)
	[CI: 2.5 – 100%]	[CI: 99.2 – 100%]
Prader-Willi	100% (15/15)	100% (424/424)
	[CI: 78.2 – 100%]	[CI: 99.1 – 100%]
Wolf-Hirschhorn	100% (2/2)	99.8% (437/438)
	[CI: 15.8 – 100%]	[CI: 98.7 – 100%]

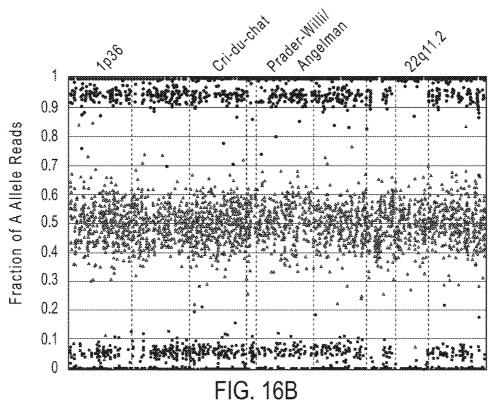
FIG. 14

U.S. Patent Dec. 20, 2022 Sheet 28 of 105 US 11,530,454 B2

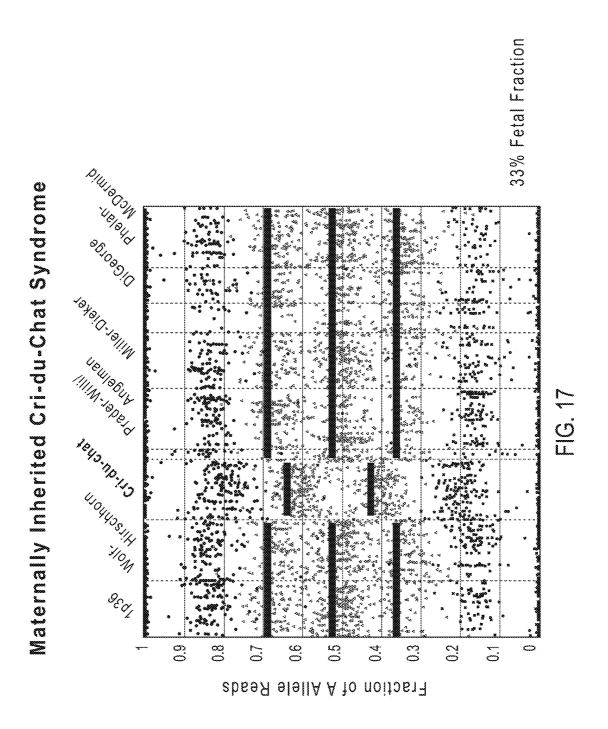


U.S. Patent Dec. 20, 2022 Sheet 29 of 105 US 11,530,454 B2

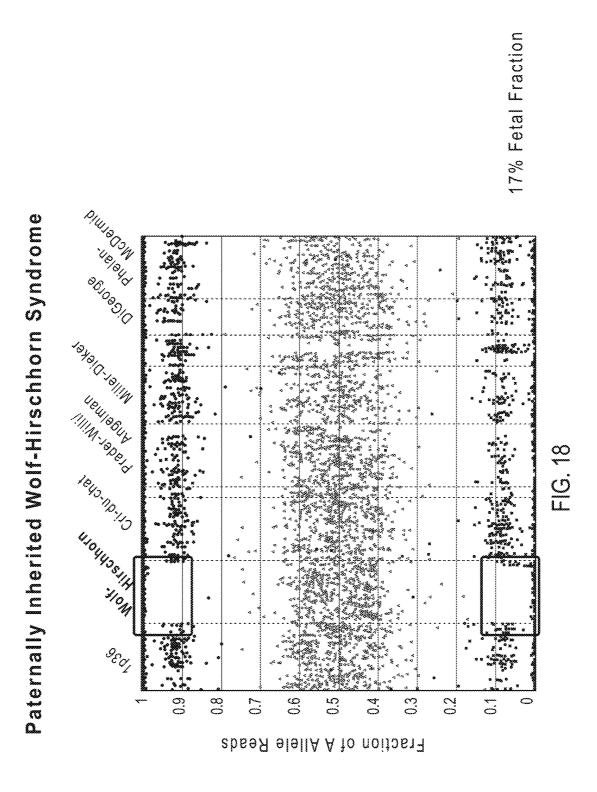


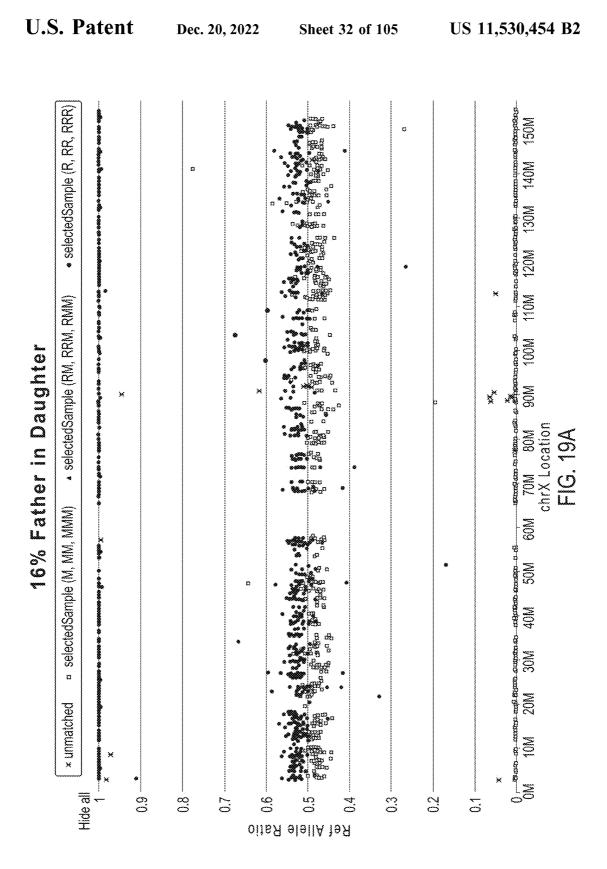


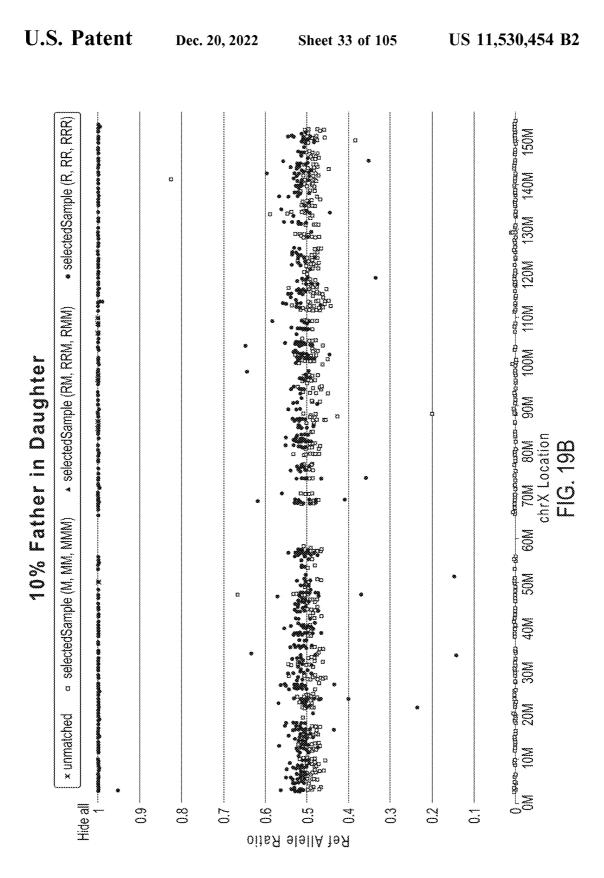
U.S. Patent Dec. 20, 2022 Sheet 30 of 105 US 11,530,454 B2

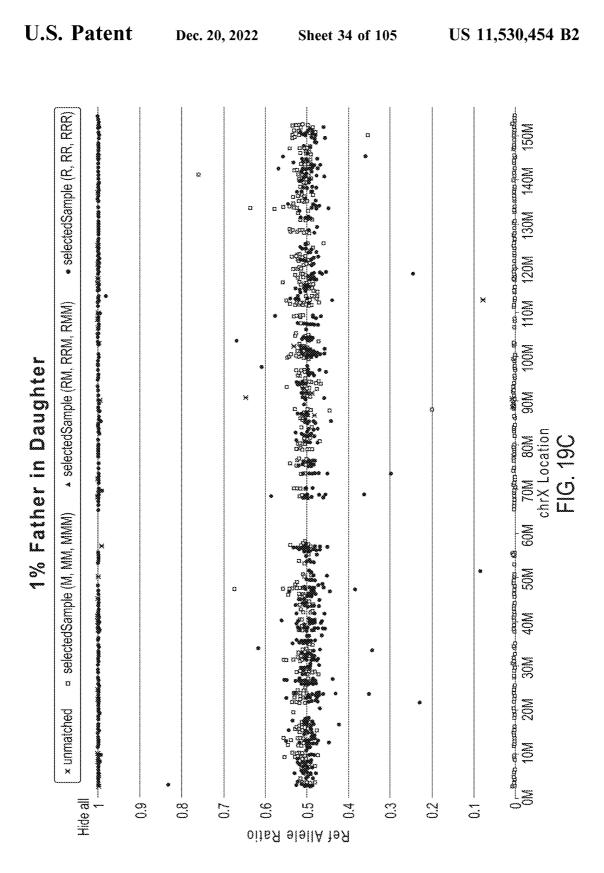


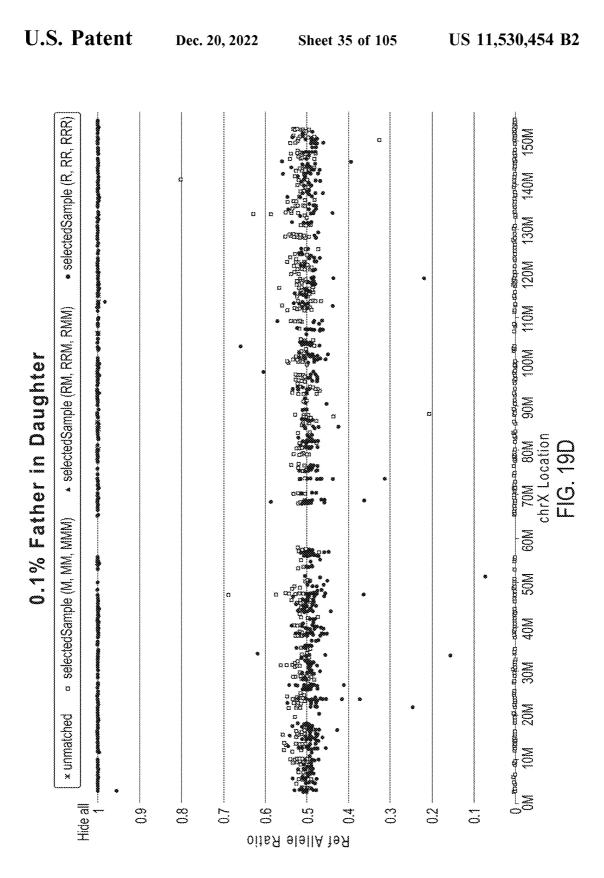
U.S. Patent Dec. 20, 2022 Sheet 31 of 105 US 11,530,454 B2



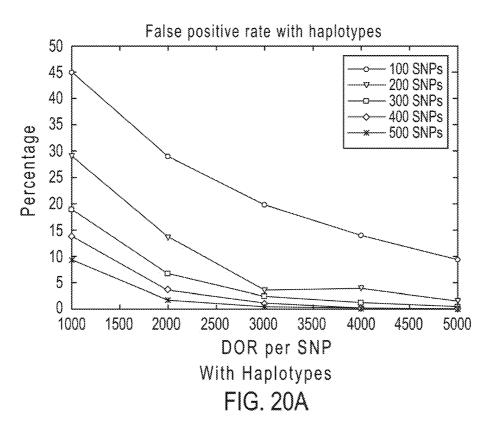


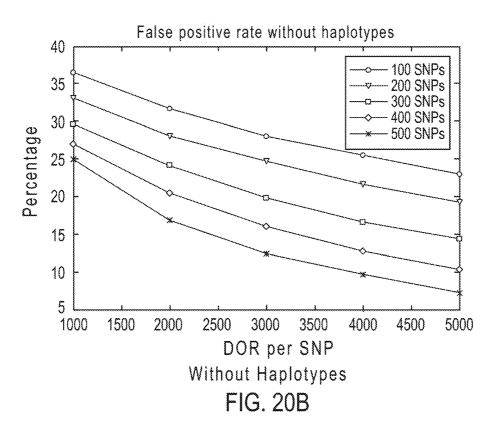




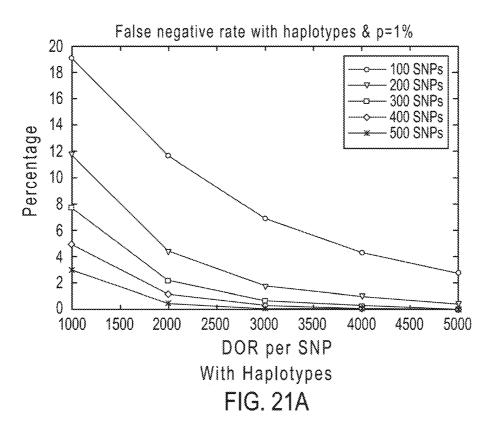


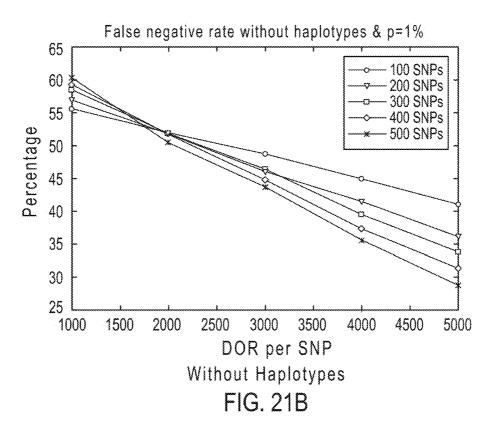
U.S. Patent Dec. 20, 2022 Sheet 36 of 105 US 11,530,454 B2



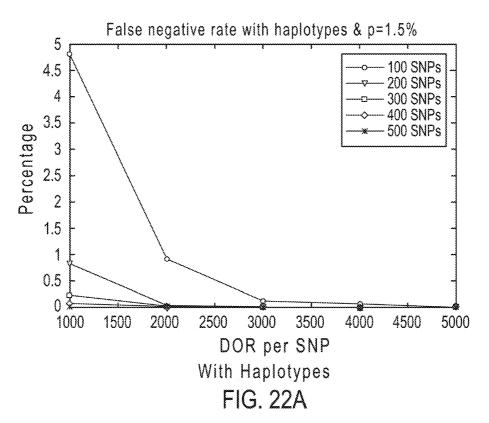


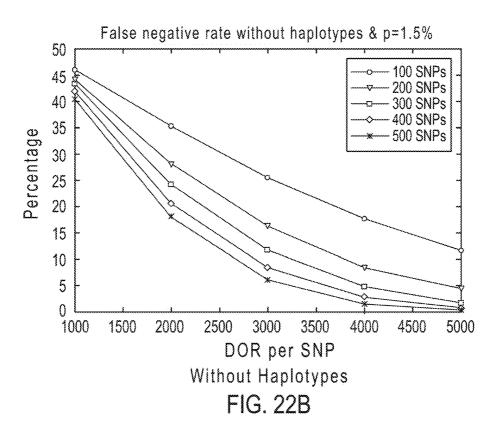
U.S. Patent Dec. 20, 2022 Sheet 37 of 105 US 11,530,454 B2



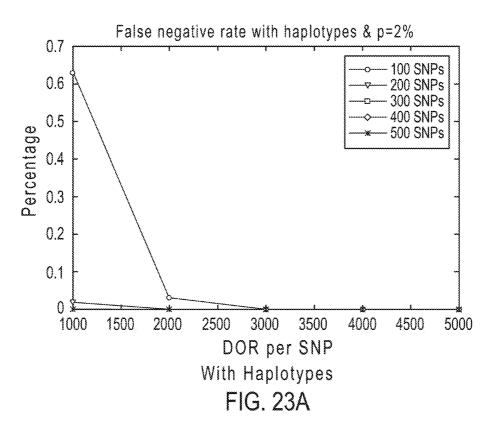


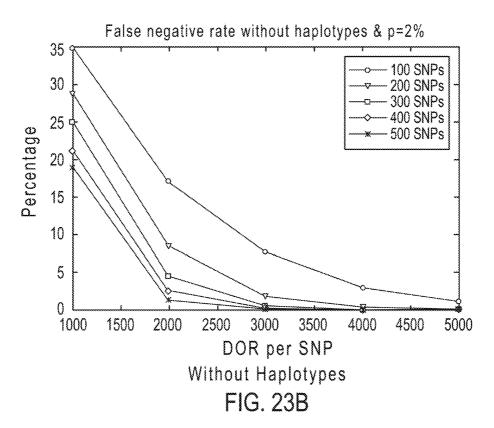
U.S. Patent Dec. 20, 2022 Sheet 38 of 105 US 11,530,454 B2



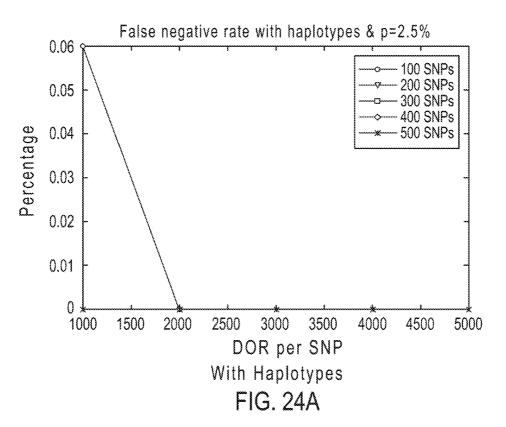


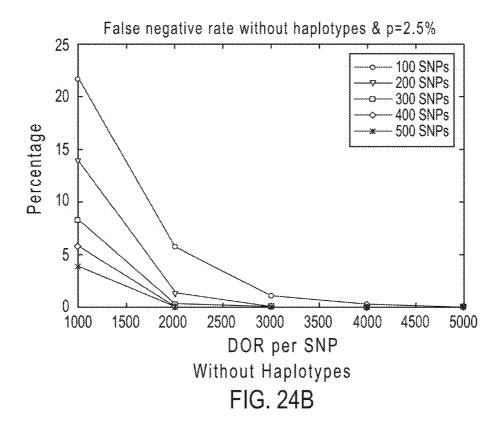
U.S. Patent Dec. 20, 2022 Sheet 39 of 105 US 11,530,454 B2



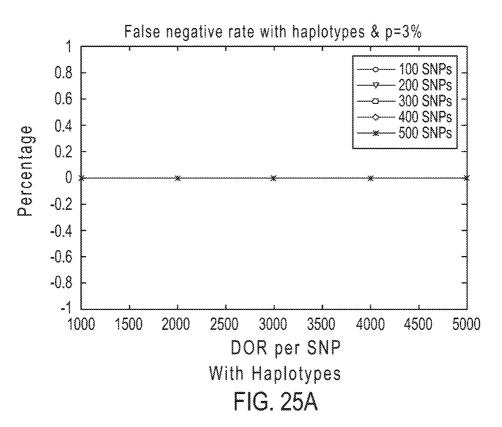


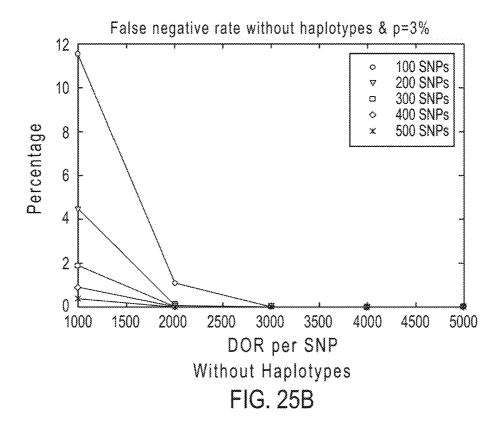
U.S. Patent Dec. 20, 2022 Sheet 40 of 105 US 11,530,454 B2





U.S. Patent Dec. 20, 2022 Sheet 41 of 105 US 11,530,454 B2

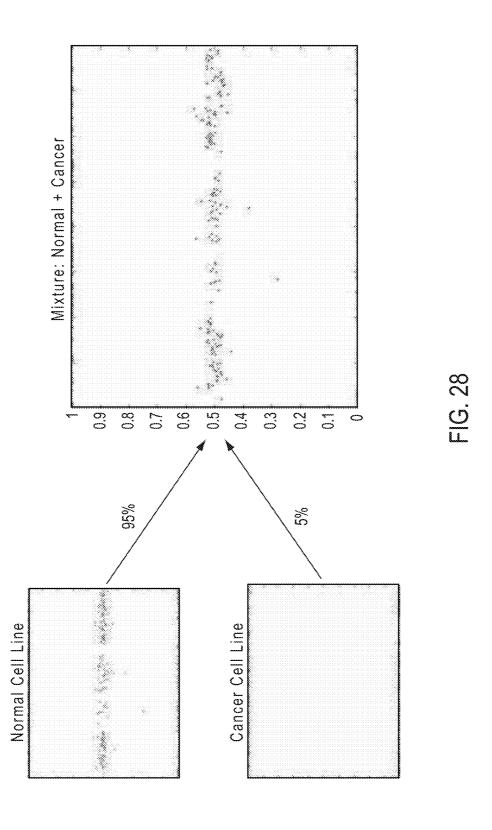




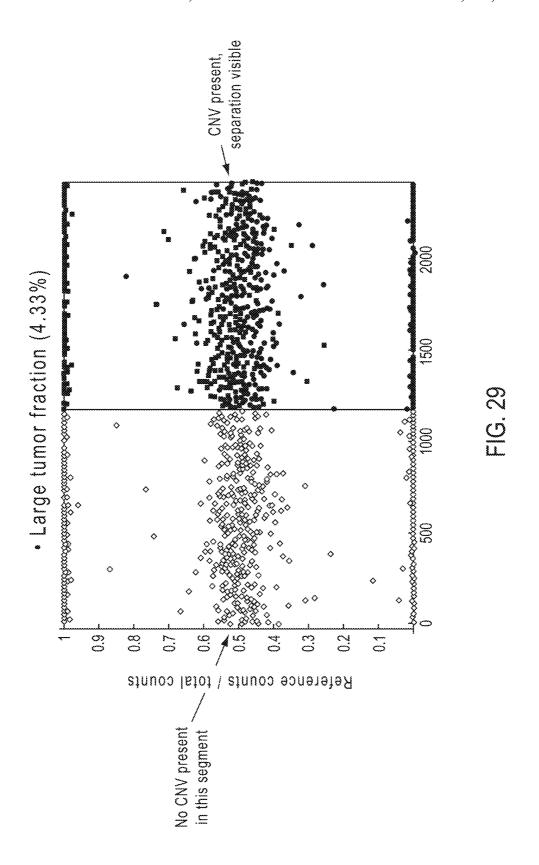
U.S. Patent Dec. 20, 2022 Sheet 42 of 105 US 11,530,454 B2

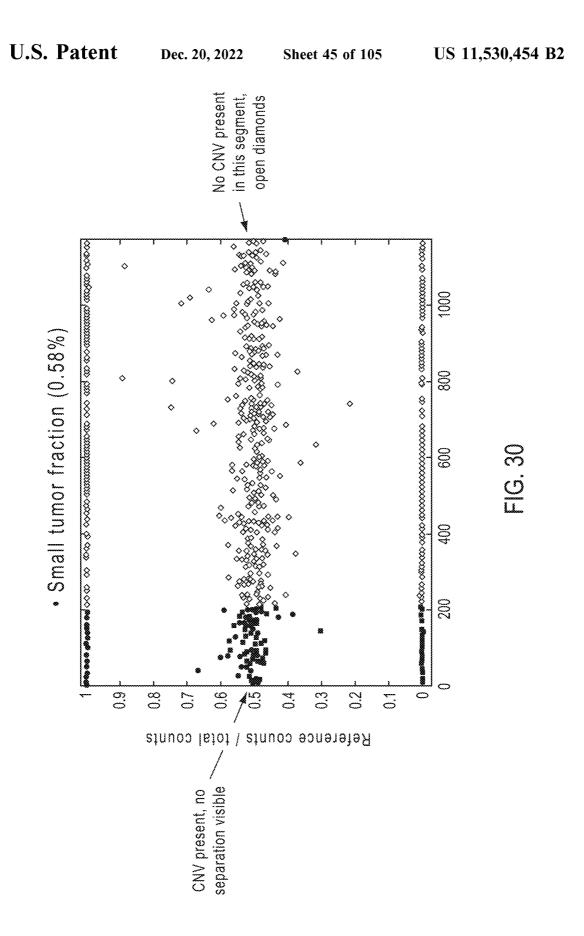
	S=1000,D=1000	S=500,D=1000	S=1000,D=500	S=500,D=500
p=1%,WithPhase	%00.0	0.10%	3.00%	9.70%
p=1%,WithoutPhase	74.20%	73.70%	74.60%	%00.92
p=2%,WithPhase	0.00%	%00.0	0.00%	%00'0
p=2%,WithoutPhase	13.90%	14.80%	41.00%	42.70%
p=3%,WithPhase	0.00%	%00.0	0.00%	%00.0
p=3%,WithoutPhase	%00.0	%00.0	5.70%	2.00%
p=4%,WithPhase	0.00%	0.00%	0.00%	0.00%
p =4%,WithoutPhase	%00.0	%00.0	0.00%	%00'0
p=5%,WithPhase	%00.0	0.00%	0.00%	%00'0
p=5%,WithoutPhase	%00'0	0.00%	0.00%	%00'0

U.S. Patent Dec. 20, 2022 Sheet 43 of 105 US 11,530,454 B2



U.S. Patent Dec. 20, 2022 Sheet 44 of 105 US 11,530,454 B2

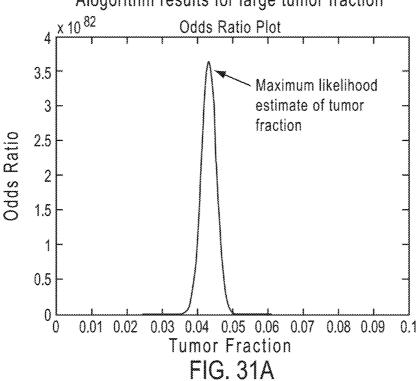




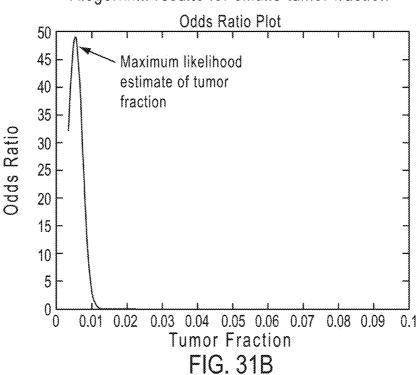
Document: 42-1 Page: 390 Case: 24-1324 Filed: 03/18/2024

U.S. Patent Dec. 20, 2022 US 11,530,454 B2 **Sheet 46 of 105** 

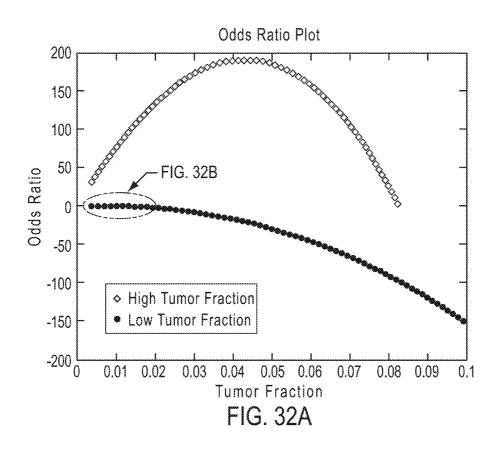
Alogorithm results for large tumor fraction

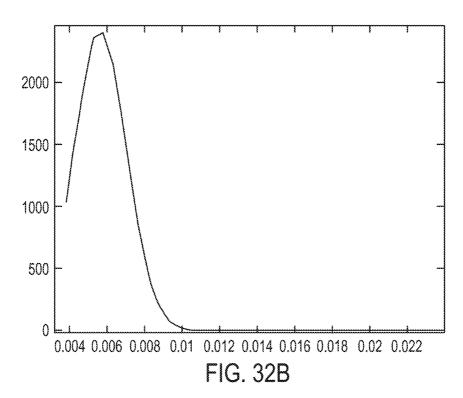


Alogorithm results for smalle tumor fraction

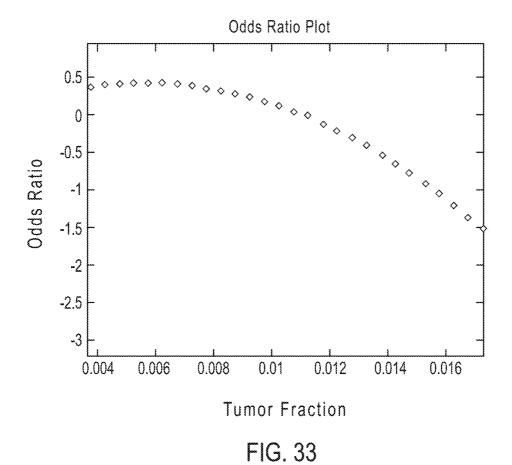


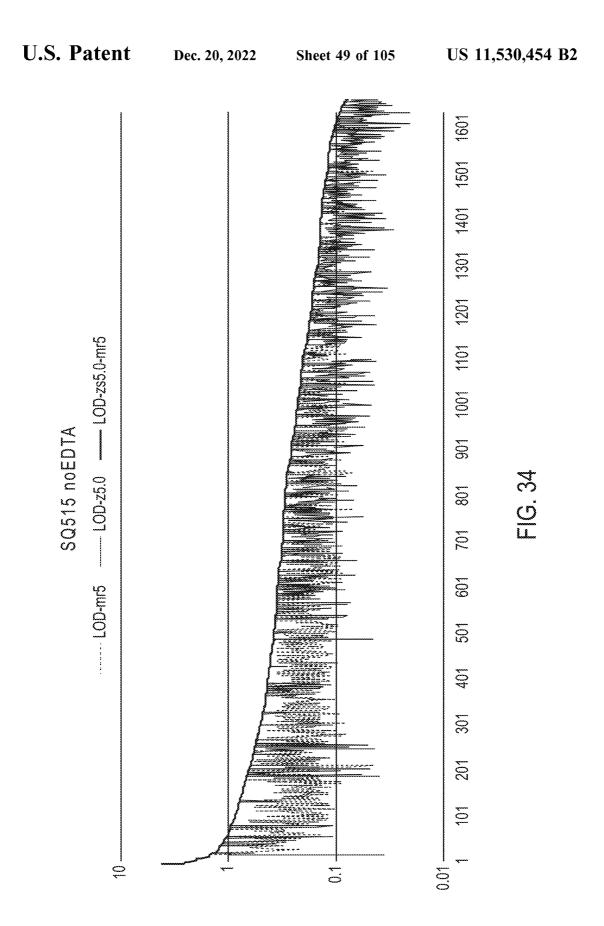
U.S. Patent Dec. 20, 2022 Sheet 47 of 105 US 11,530,454 B2

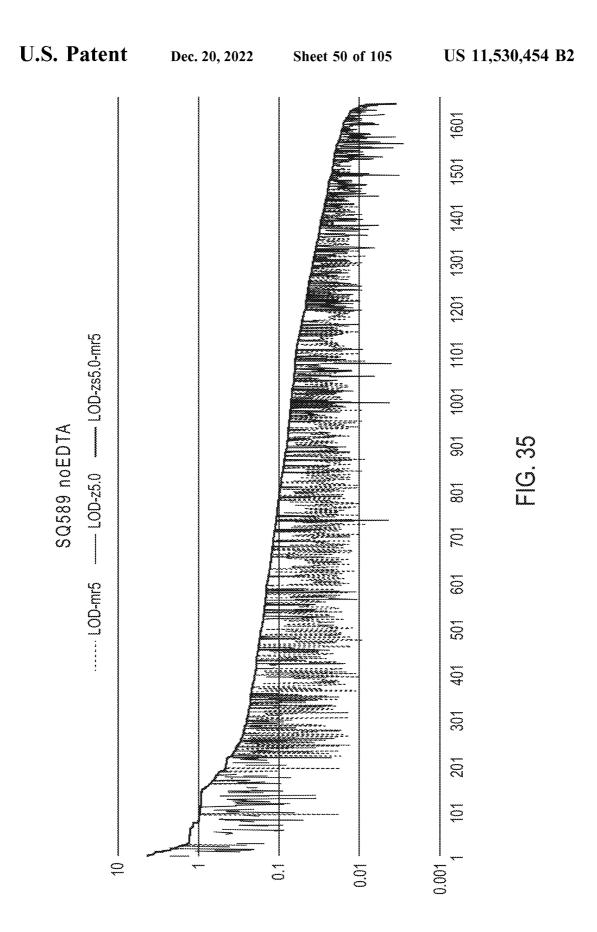




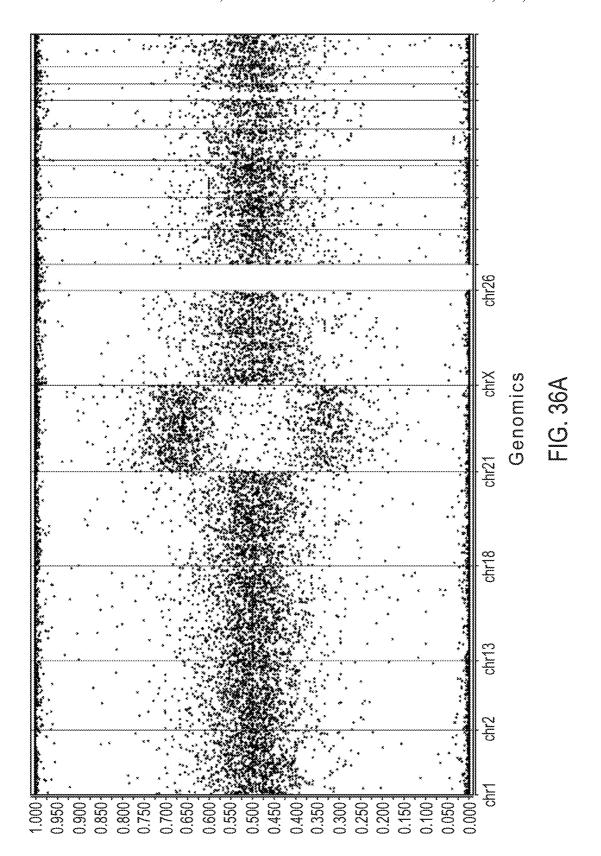
U.S. Patent Dec. 20, 2022 Sheet 48 of 105 US 11,530,454 B2



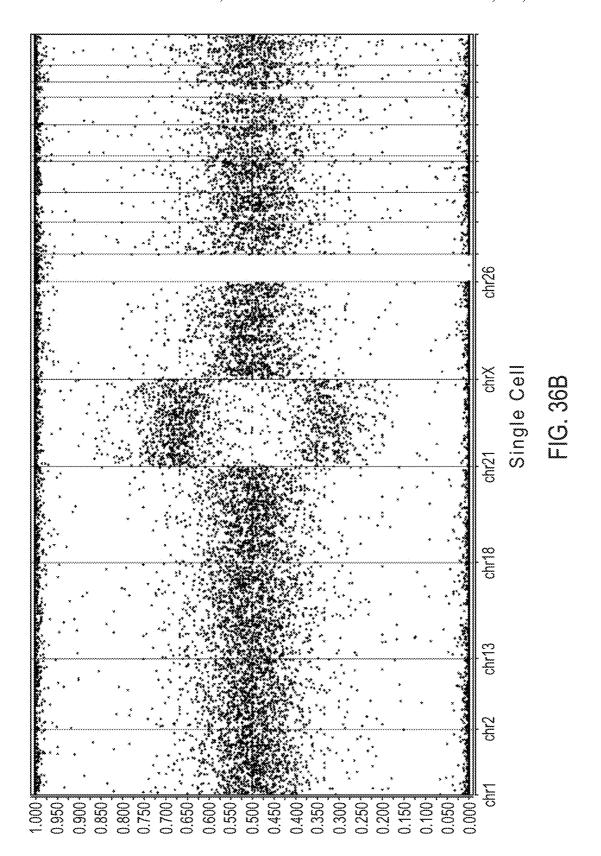




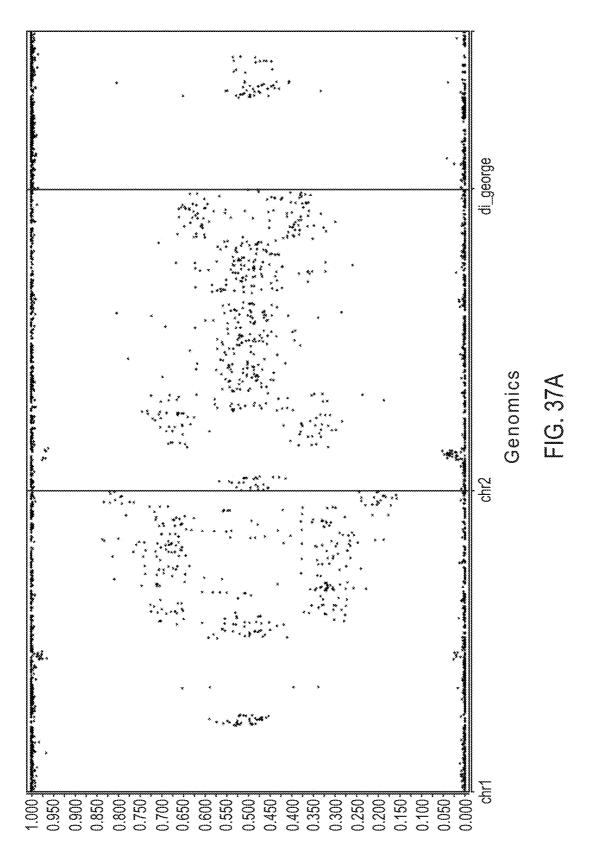
U.S. Patent Dec. 20, 2022 Sheet 51 of 105 US 11,530,454 B2



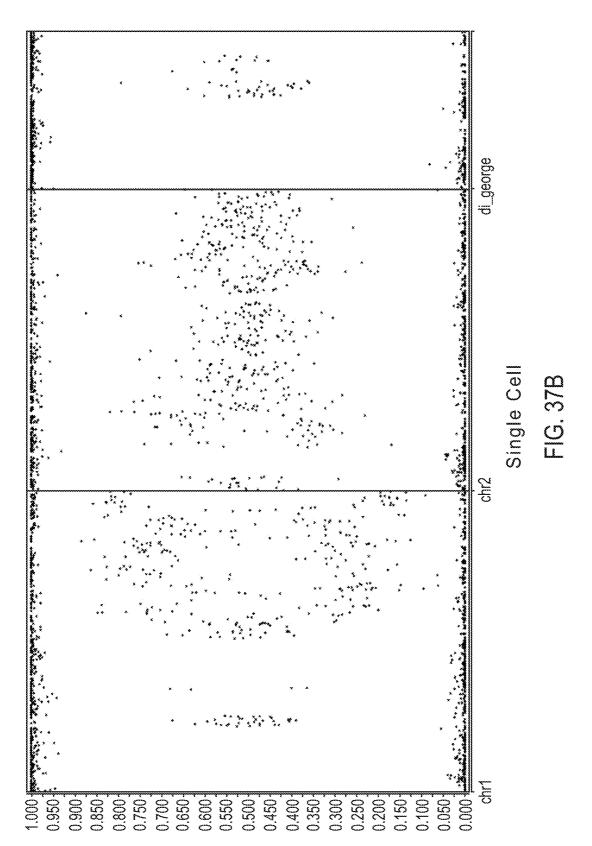
U.S. Patent Dec. 20, 2022 Sheet 52 of 105 US 11,530,454 B2



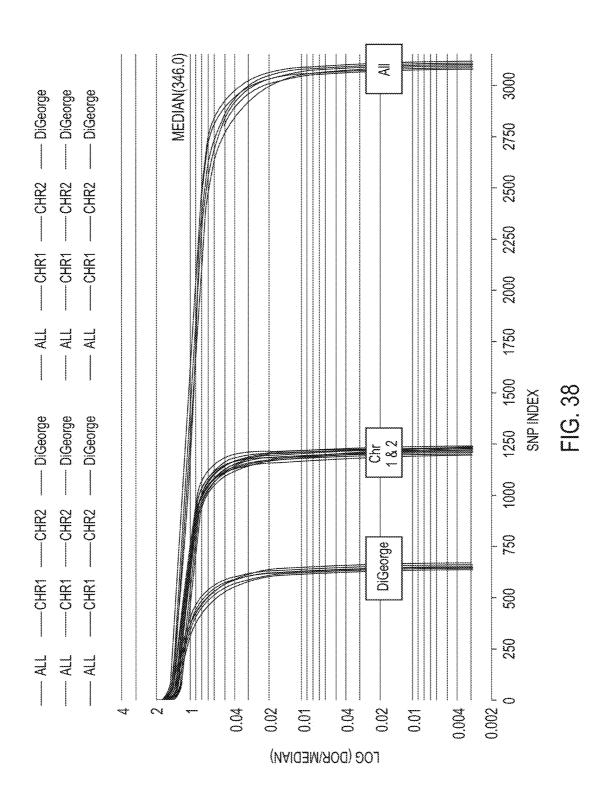
U.S. Patent Dec. 20, 2022 Sheet 53 of 105 US 11,530,454 B2



U.S. Patent Dec. 20, 2022 Sheet 54 of 105 US 11,530,454 B2



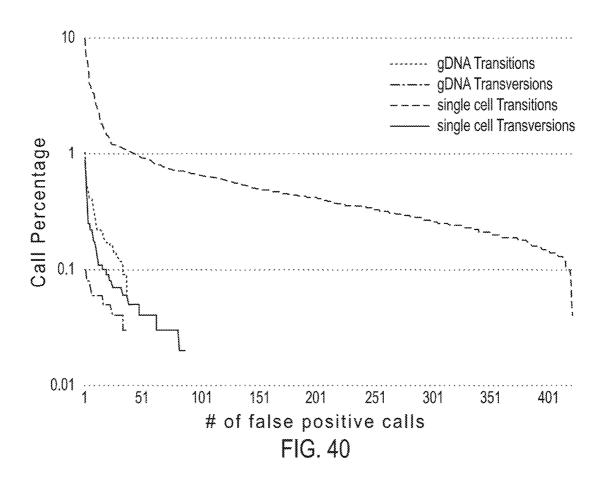
U.S. Patent Dec. 20, 2022 Sheet 55 of 105 US 11,530,454 B2



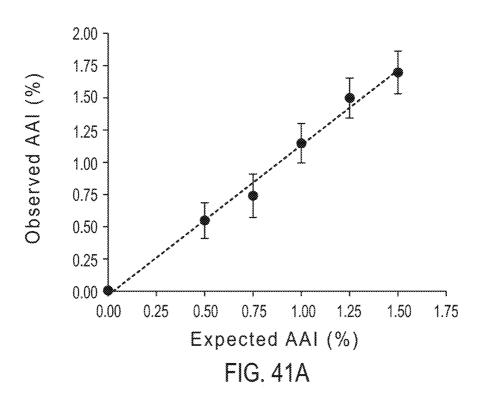
U.S. Patent Dec. 20, 2022 Sheet 56 of 105 US 11,530,454 B2

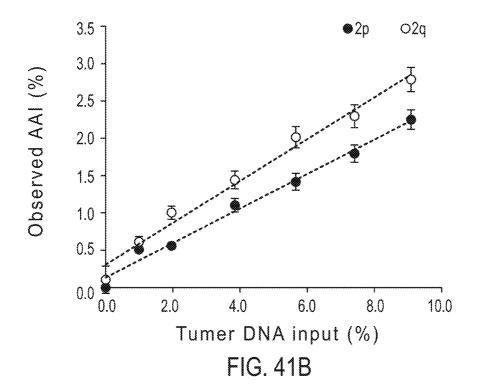
	gDNA	Single Cell
Count	75	510
Mean	0.15 %	0.51 %
Median	0.09 %	0.33 %
Max	1.03 %	10 %
Standard Deviation	0.16%	0.79 %
95 <sup>th</sup> percentile	0.43 %	1.22 %
90 <sup>th</sup> percentile	0.37 %	0.92 %

FIG. 39

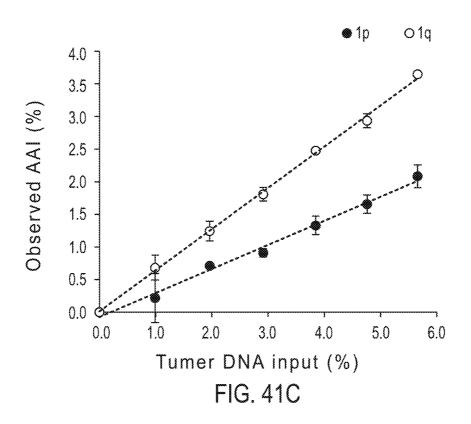


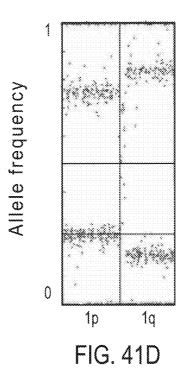
U.S. Patent Dec. 20, 2022 Sheet 57 of 105 US 11,530,454 B2



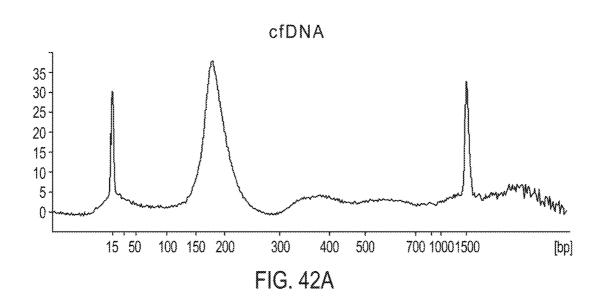


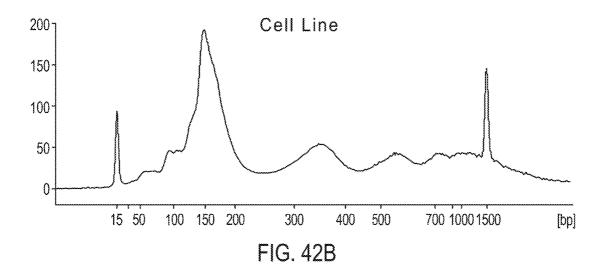
U.S. Patent Dec. 20, 2022 Sheet 58 of 105 US 11,530,454 B2



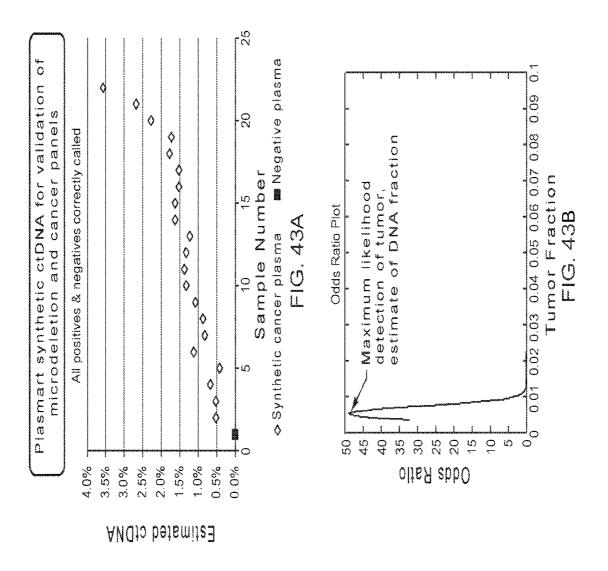


U.S. Patent Dec. 20, 2022 Sheet 59 of 105 US 11,530,454 B2

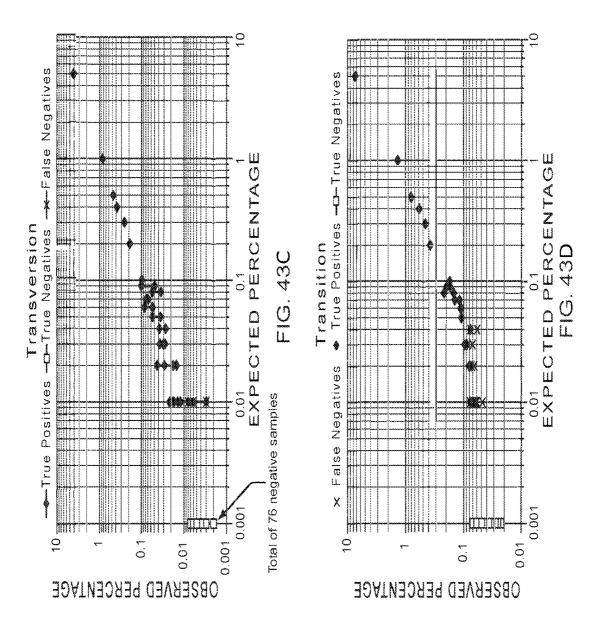


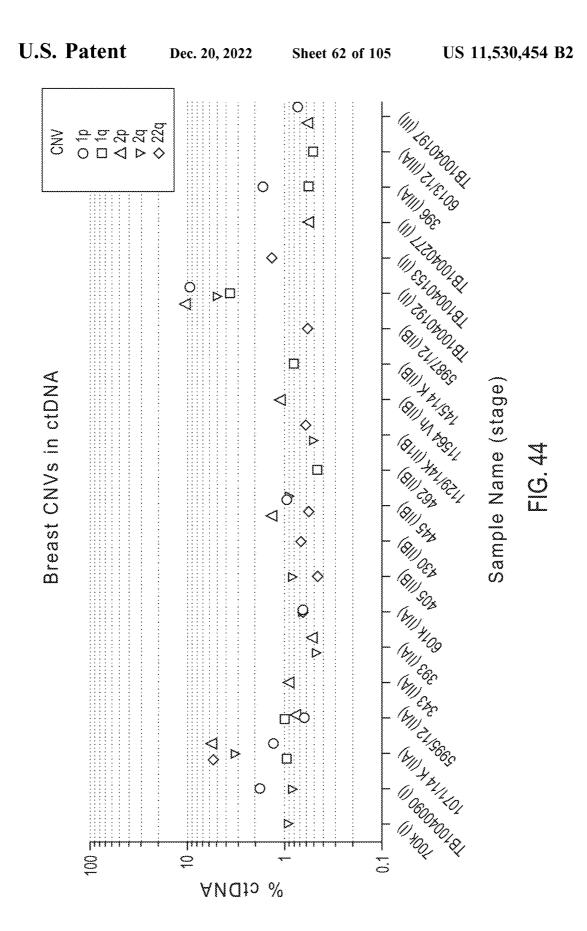


U.S. Patent Dec. 20, 2022 Sheet 60 of 105 US 11,530,454 B2

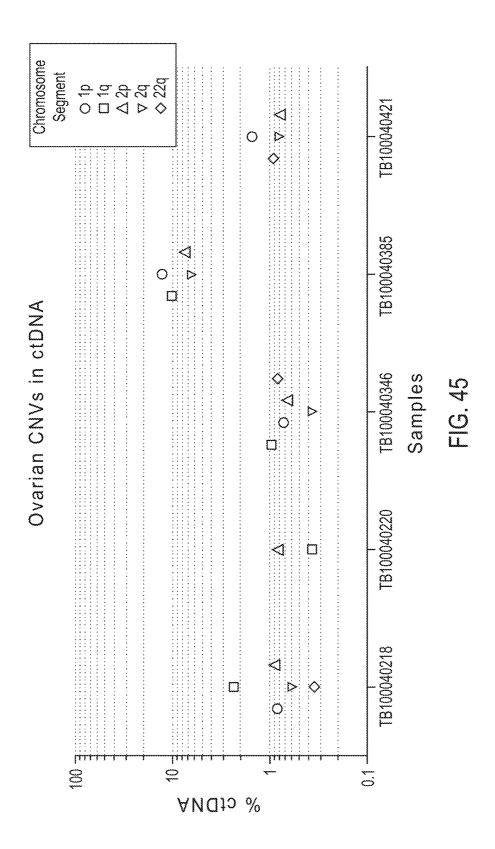


U.S. Patent Dec. 20, 2022 Sheet 61 of 105 US 11,530,454 B2





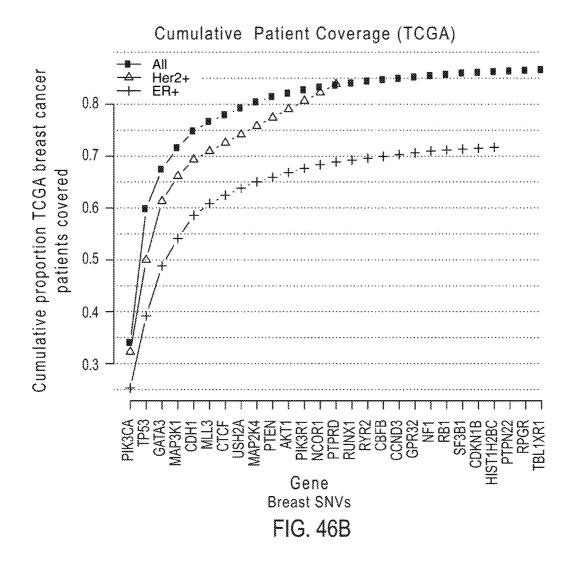
U.S. Patent Dec. 20, 2022 Sheet 63 of 105 US 11,530,454 B2



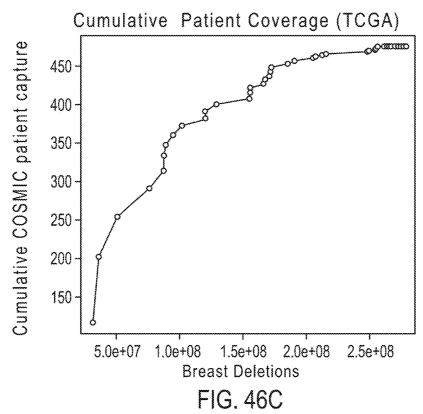
U.S. Patent Dec. 20, 2022 Sheet 64 of 105 US 11,530,454 B2

		SNVs	Only	Combined SNVs & CNVs		
Cancer Type	Total # of Positives	Detected	%	Detected w/ low Depth of Read	%	Expected Diagnostic Load Based on TCGA and COMSIC Datasets
Breast	41	29	71%	34	83%	>97%
Lung	24	17	71%	22	92%	>98%

FIG. 46A



U.S. Patent Dec. 20, 2022 Sheet 65 of 105 US 11,530,454 B2



Cumulative Patient Coverage (TCGA)

900

0e+00

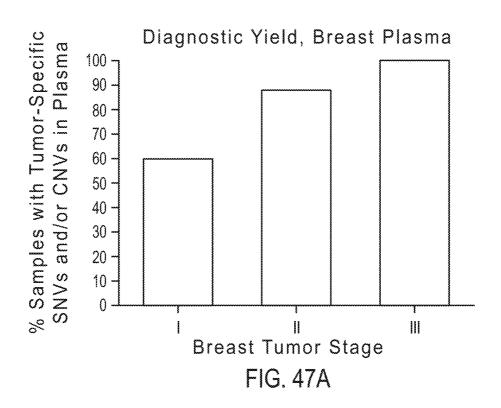
1e+07

2e+07

Breast Amplifications

FIG. 46D

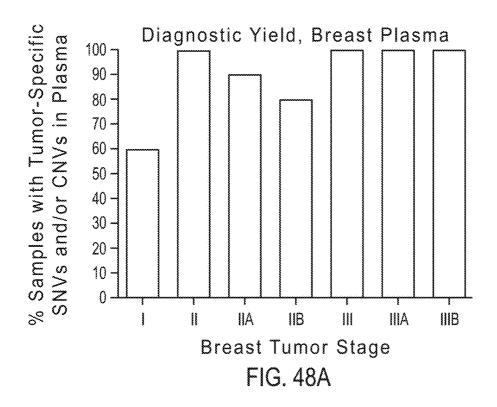
U.S. Patent Dec. 20, 2022 Sheet 66 of 105 US 11,530,454 B2



Stage	Positive Samples, N	Detected, N	% Detected	
l	10	6	60%	
anns	25	22	88%	
III	6	6	100%	

LOQ for SNVs= 0.2% ctDNA LOQ for CNVs= 0.45% ctDNA FIG. 47B

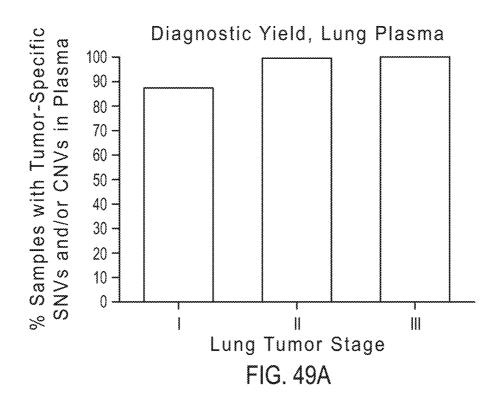
U.S. Patent Dec. 20, 2022 Sheet 67 of 105 US 11,530,454 B2



Stage	Positive Samples, N	Detected, N	% Detected	
l	10	6	60%	
II	5	5	100%	
IIA	10	9	90%	
IIB	10	8	80%	
III	3	3	100%	
IIIA	2	2	100%	
IIIB	1	1	100%	

LOQ for SNVs= 0.2% ctDNA LOQ for CNVs= 0.45% ctDNA FIG. 48B

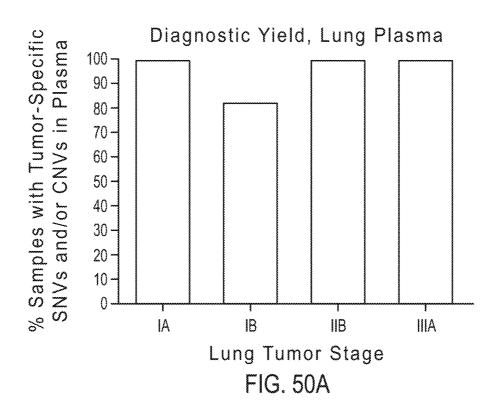
U.S. Patent Dec. 20, 2022 Sheet 68 of 105 US 11,530,454 B2



Stage	Positive Samples, N	Detected, N	% Detected	
I	17	15	88%	
II	5	5	100%	
III 2		2	100%	

LOQ for SNVs= 0.2% ctDNA LOQ for CNVs= 0.45% ctDNA FIG. 49B

U.S. Patent Dec. 20, 2022 Sheet 69 of 105 US 11,530,454 B2



Stage	Positive Samples, N	Detected, N	% Detected	
IA	6	6	100%	
IB	IB 11		82%	
IIB	5	5	100%	
IIIA	2	2	100%	

LOQ for SNVs= 0.2% ctDNA LOQ for CNVs= 0.45% ctDNA FIG. 50B

U.S. Patent Dec. 20, 2022 Sheet 70 of 105 US 11,530,454 B2

	Histology	Stage	Age	Dia (mm)	Smoker (Pack Years)	
L12	SSC	IB	69	40	Yes (?)	
L13	SSC	IA	68	30	Yes (100)	
L15	SSC	IB	68	50	Yes (50)	
L17	Adeno	IIB	61	20	Yes (48)	

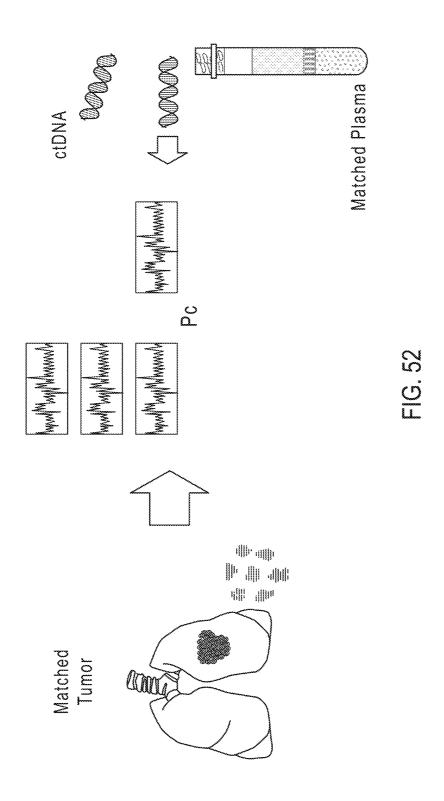
FIG. 51A

U.S. Patent Dec. 20, 2022 Sheet 71 of 105 US 11,530,454 B2

Sample	Gene	Chr	ChrStart	VAF R1	VAF R2	VAF R3
L12	BRIP1	chr17	59924572	14	6	8
L12	CARS	chr11	3062181	22	11	16
L12	CIC	chr19	42797381	0	0	7
L12	CYFIP1	chr15	22940733	0	6	O
L12	FAT1	chr4	187519147	8	4	0
L12	KDM6A	chrX	44921898	10	5	0
L12	MLLT4	chr6	168347475	9	3	0
L12	NFE2L2	chr2	178098801	69	31	50
L12	RASA1	chr5	86642517	7	0	O
L12	TP53	chr17	7578406	23	9	17
L12	TP53	chr17	7578190	22	8	16
L13	EGFR	chr7	55241708	21	25	55
L13	EGFR	chr7	55242511	20	17	48
L13	HERC4	chr10	69793756	0	3	6
L13	JAK2	chr9	5022084	0	11	7
L13	KMT2C	chr7	151947008	0	0	4
L13	MSH2	chr2	47693816	5	0	11
L13	MTOR	chr1	11292495	2	0	3
L13	PLCG2	chr16	81942036	5	0	6
L13	TP53	chr17	7579509	7	4	16
L15	ALK	chr2	29940530	6	2	
L15	GABRG1	chr4	46060315	13	0	
L15	KDM6A	chrX	44922755	18	6	
L15	MLL2	chr12	49443815	5	0	
L15	ROS1	chr6	117687379	10	L	
L15	SLC39A4	chr8	145638322	14		
L15	TP53	chr17	7578254	19	3	
L15	ZFHX4	chr8	77776735	13		
L15		<u> </u>	<b></b>			
FT2	ZMYM4	chr1	35827319	14	<u> </u>	<u> </u>
L17	BRCA2	chr13	32914959	16	28	0
L17	KRAS	chr12	25398284	13	22	0
L17	NF1	chr17	29653134	20	25	0
L17	NF1	chr17	29528088	0	0	20
L17	PAX8	chr2	113984793	9	0	0
L17	TP53	chr17	7577610	0	0	16
L17	TP53	chr17	7577079	20	35	0
L17	TRIM67	chr1	231299607	0	50	0
L17	TRIP11	chr14	92471631	7	0	0

FIG. 51B

U.S. Patent Dec. 20, 2022 Sheet 72 of 105 US 11,530,454 B2



U.S. Patent Dec. 20, 2022 Sheet 73 of 105 US 11,530,454 B2

Sample	Gene	Chrom	ChromStart	Swanton			
				VAF R1	VAF R2	VAF R3	
L12	BRIP1	chr17	59924572	14	6	8	
L12	CARS	chr11	3062181	22	11	16	
L12	CIC	chr19	42797381	0	0	7	
L12	CYFIP1	chr15	22940733	0	///// 6	0	
L12	FAT1	chr4	187519147	8	4	/////// 0	
L12	KDM6A	chrX	44921898	10	5	0	
L12	MLLT4	chr6	168347475	9	3	////// 0	
L12	NFE2L2	chr2	178098801	69	31	50	
L12	RASA1	chr5	86642517	7	0	/////// 0	
L12	TP53	chr17	7578406	23	9	17	
L12	TP53	chr17	7578190	22	8	16	
L13	EGFR	chr7	55241708	21	25	55	
L13	EGFR	chr7	55242511	20	17	48	
L13	HERC4	chr10	69793756	///// O	3	6	
L13	JAK2	chr9	5022084	0	11	7	
L13	KMT2C	chr7	151947008	0	0	4	
L13	MSH2	chr2	47693816	5	////// ol	11	
L13	MTOR	chr1	11292495	2	ol	3	
L13	PLCG2	chr16	81942036	5	///// o	6	
L13	TP53	chr17	7579509	7	4	16	
L15	ALK	chr2	29940530	6	2	dinici di	
L15	GABRG1	chr4	46060315	13	////// o		
L15	KDM6A	chrX	44922755	18	6		
L15	MLL2	chr12	49443815	5	0		
L15	ROS1	chr6	117687379	10	3		
L15	SLC39A4	chr8	145638322	14	0	••••••	
L15	TP53	chr17	7578254	19	3		
L15	ZFHX4	chr8	77776735	13	5	······	
L15	ZMYM4	chr1	35827319	14	2	•••••	
L17	BRCA2	chr13	32914959	16		***************************************	
L17	KRAS	chr12	25398284	13	22		
L17	NF1	chr17	29653134	20	25	***************************************	
L17	NF1	chr17	29528088	0	0	20	
L17	PAX8	chr2	113984793	9	0	0	
L17	TP53	chr17	7577610	0	Ō	16	
L17	TP53	chr17	7577079	20	35	0	
L17	TRIM67	chr1	231299607	////// 0	50	0	
L17	TRIP11	chr14	92471631	7	0	0	

= Ampliseq

FIG. 53A

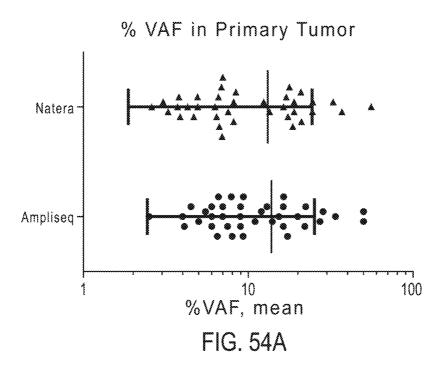
U.S. Patent Dec. 20, 2022 Sheet 74 of 105 US 11,530,454 B2

Sample	Gene	Chrom	ChromStart		Na	tera	
				VAF R1	VAF R2	VAF R3	VAF Plasma
L12	BRIP1	chr17	59924572	22.6	7.74	10.07	3.71
L12	CARS	chr11	3062181	22.39	10.47	18.21	5.02
L12	CIC	chr19	42797381	0	0	8.17	1.71
L12	CYFIP1	chr15	22940733	0	////// o	0	0
L12	FAT1	chr4	187519147	9.81	4.26	0.77	0.57
L12	KDM6A	chrX	44921898	6.33	3.54	0	0.28
L12	MLLT4	chr6	168347475	9.04	4.26	0.73	0.95
L12	NFE2L2	chr2	178098801	73.12	39.92	53.94	23.25
L12	RASA1	chr5	86642517	7.37	0	0.35	0
L12	TP53	chr17	7578406	22.93	9.17	16.97	4.89
L12	TP53	chr17	7578190	21.48	8.67	22.09	5,77
L13	EGFR	chr7	55241708	27.06	23.14	60.79	1.16
L13	EGFR	chr7	55242511	23.93	18.01	56.31	1.09
L13	HERC4	chr10	69793756	0.16	1.92	5.72	0
L13	JAK2	chr9	5022084	0.27	2.79	6.09	0
L13	KMT2C	chr7	151947008	0	0	6.69	0
L13	MSH2	chr2	47693816	5.04	3.33	12.67	0
L13	MTOR	chr1	11292495	2.4	1.38	6.01	0
L13	PLCG2	chr16	81942036	2.8	1.7	6.89	0
L13	TP53	chr17	7579509	6.11	3.76	15.4	0.4
L15	ALK	chr2	29940530	7.2	1.39		0
L15	GABRG1	chr4	46060315	11.84	2.04		0
L15	KDM6A	chrX	44922755	6.54	0.91		0.17
L15	MLL2	chr12	49443815	12.41	0		
L15	ROS1	chr6	117687379	10.82	2.94		0.15
L15	SLC39A4	chr8	145638322	18.59	0		0
L15	TP53	chr17	7578254	12.6	2.5		0
L15	ZFHX4	chr8	77776735	10.26	2.3		0
L15	ZMYM4	chr1	35827319	12.84	3.46		0
L17	BRCA2	chr13	32914959	15.03	27	***************************************	0
L17	KRAS	chr12	25398284	14.22	23.79		0
L17	NF1	chr17	29653134	18.6	23.1		0
L17	NF1	chr17	29528088	0	0	17.79	0.37
L17	PAX8	chr2	113984793	6.6	0		0
L17	TP53	chr17	7577610	0	0	18.97	0
L17	TP53	chr17	7577079	15.36	33.7		0
L17	TRIM67	chr1	231299607	18.38	30.95		0.57
L17	RIP11	chr14	92471631	6.44	0		0

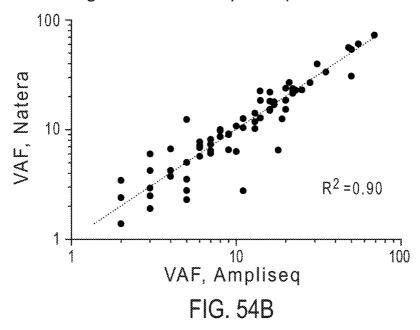
= mmPCR-NGS

FIG. 53B

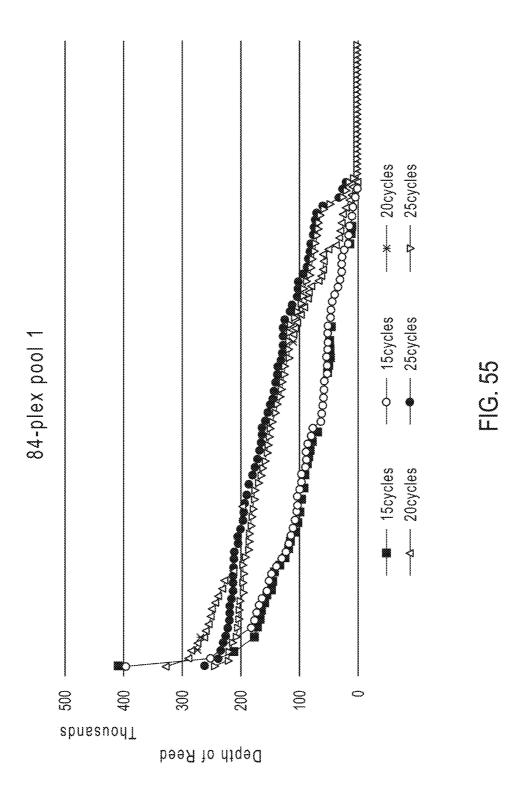
U.S. Patent Dec. 20, 2022 Sheet 75 of 105 US 11,530,454 B2



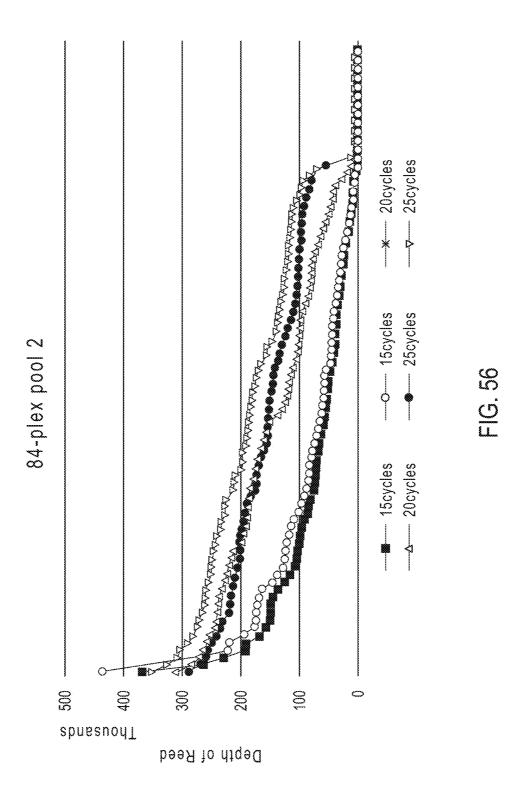
Linear Regression of AmpliSeq VAF v. Natera VAF



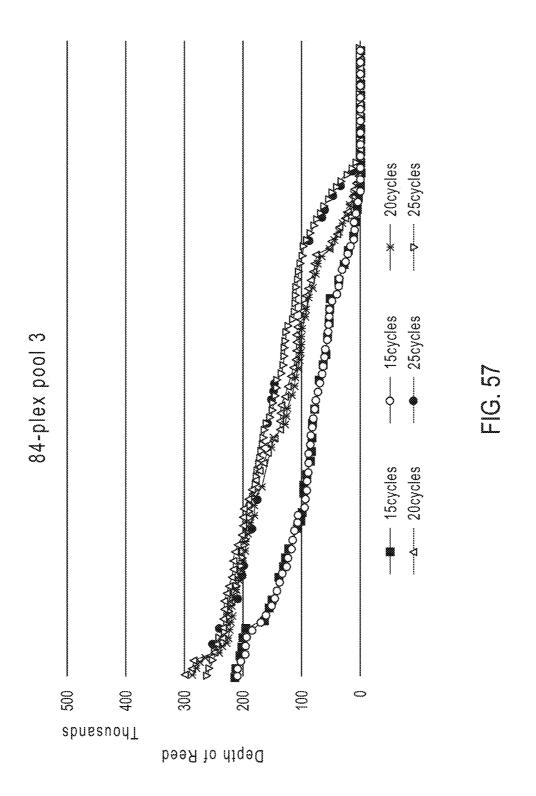
U.S. Patent Dec. 20, 2022 Sheet 76 of 105 US 11,530,454 B2



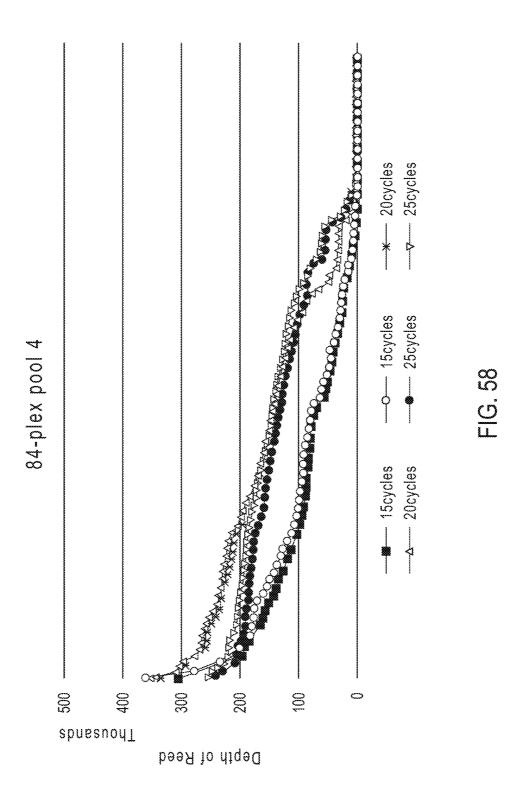
U.S. Patent Dec. 20, 2022 Sheet 77 of 105 US 11,530,454 B2

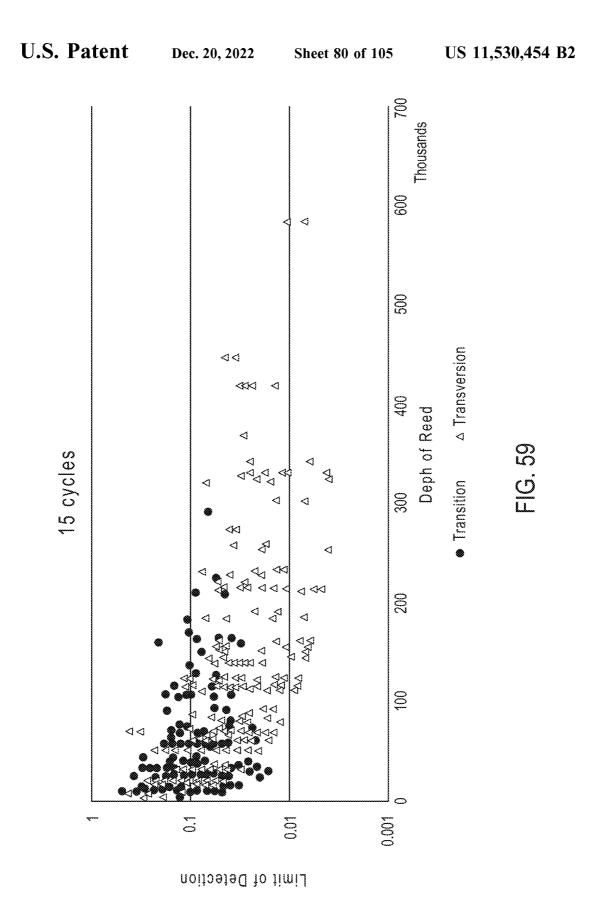


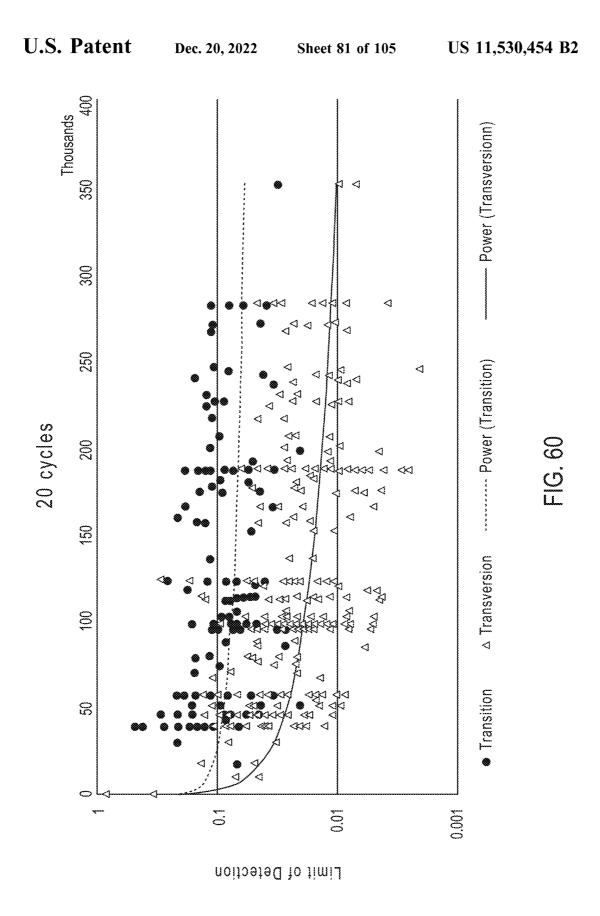
U.S. Patent Dec. 20, 2022 Sheet 78 of 105 US 11,530,454 B2

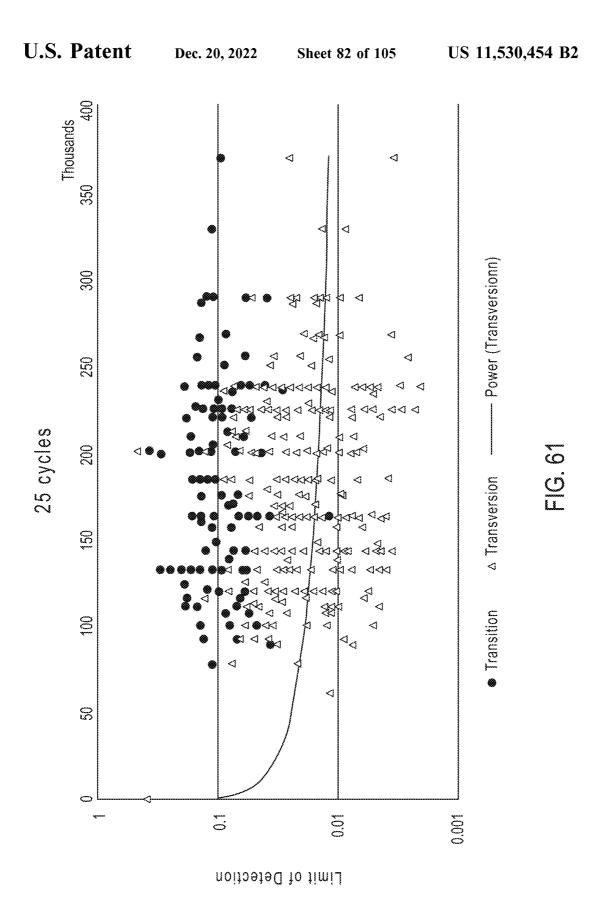


U.S. Patent Dec. 20, 2022 Sheet 79 of 105 US 11,530,454 B2

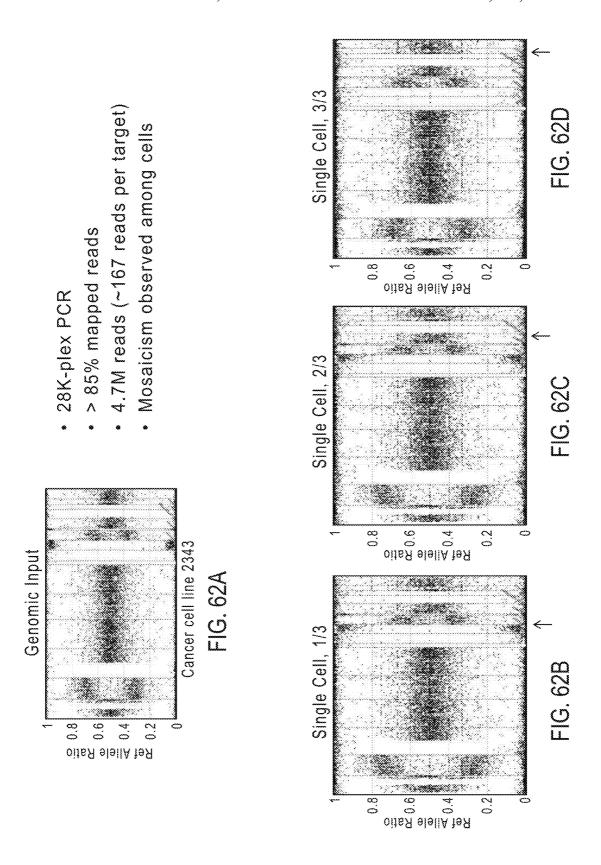








U.S. Patent Dec. 20, 2022 Sheet 83 of 105 US 11,530,454 B2



U.S. Patent Dec. 20, 2022 Sheet 84 of 105 US 11,530,454 B2

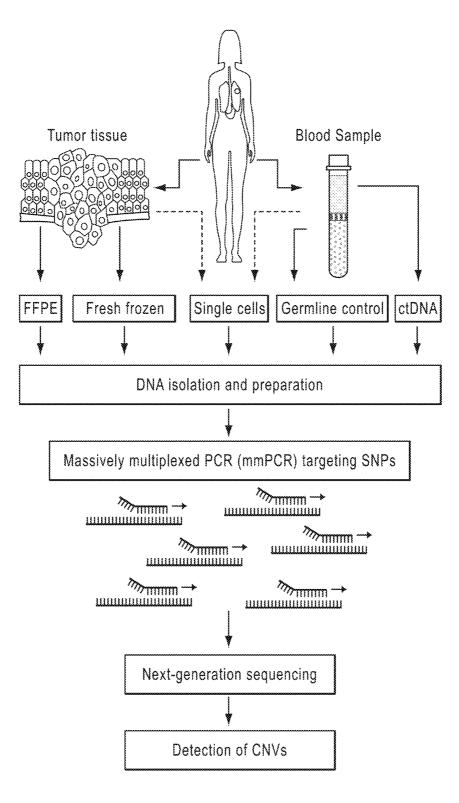


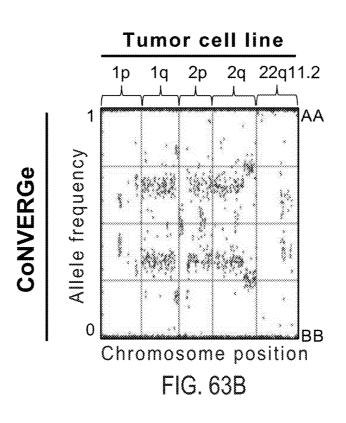
FIG. 63A

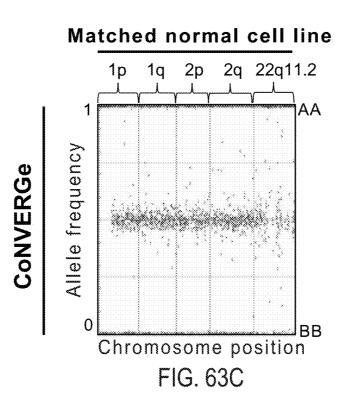
U.S. Patent

Dec. 20, 2022

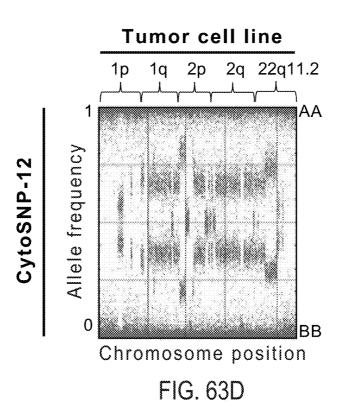
**Sheet 85 of 105** 

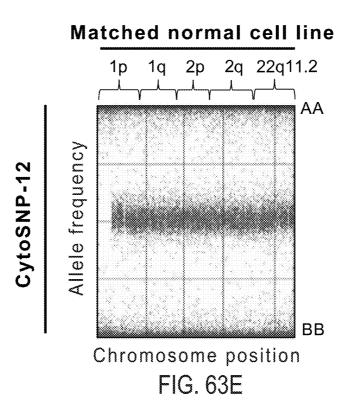
US 11,530,454 B2



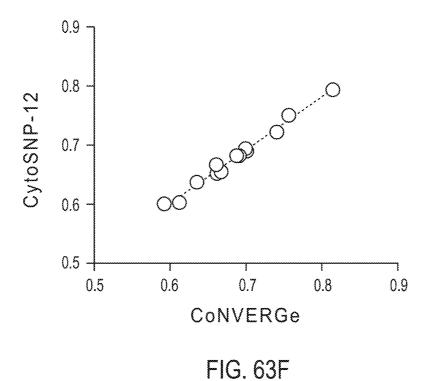


U.S. Patent Dec. 20, 2022 Sheet 86 of 105 US 11,530,454 B2

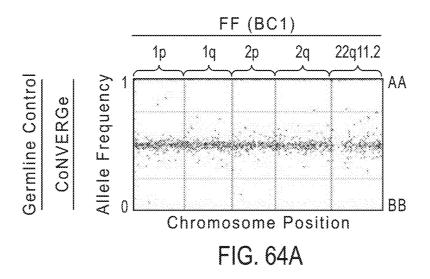




U.S. Patent Dec. 20, 2022 Sheet 87 of 105 US 11,530,454 B2



U.S. Patent Dec. 20, 2022 Sheet 88 of 105 US 11,530,454 B2



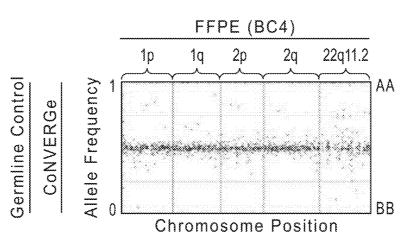
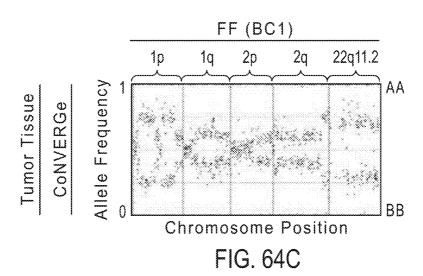
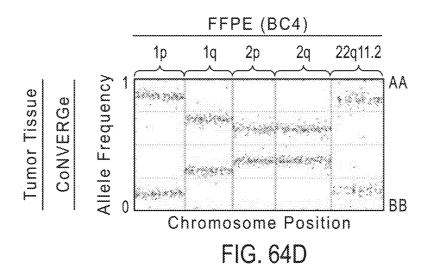
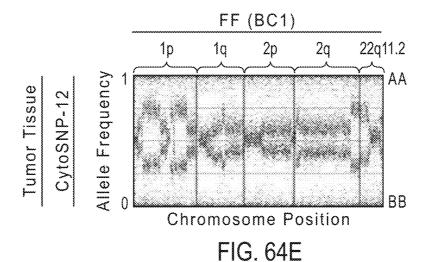


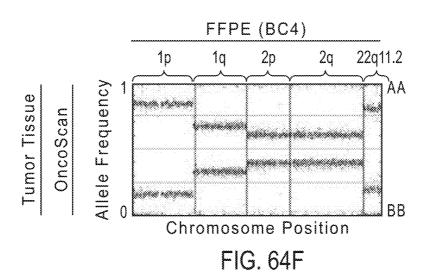
FIG. 64B



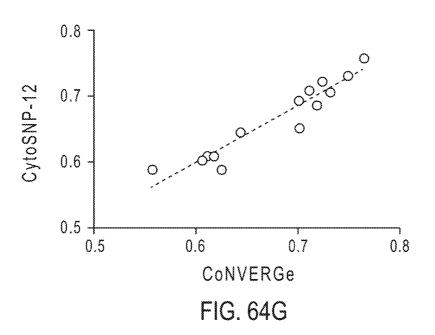
U.S. Patent Dec. 20, 2022 Sheet 89 of 105 US 11,530,454 B2

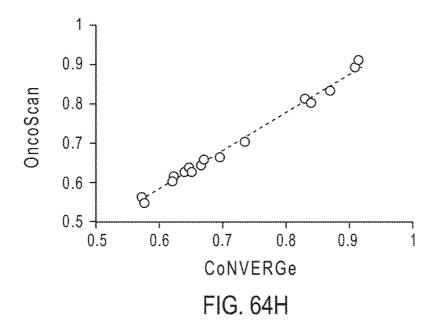






U.S. Patent Dec. 20, 2022 Sheet 90 of 105 US 11,530,454 B2





U.S. Patent Dec. 20, 2022 Sheet 91 of 105 US 11,530,454 B2

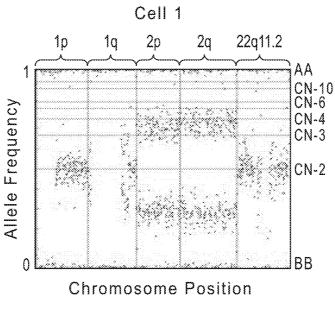
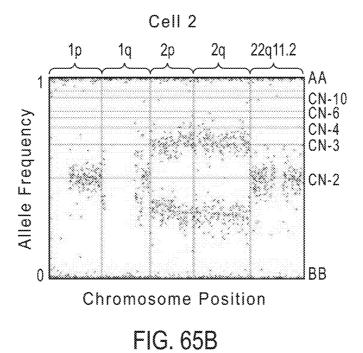


FIG. 65A



U.S. Patent Dec. 20, 2022 Sheet 92 of 105 US 11,530,454 B2

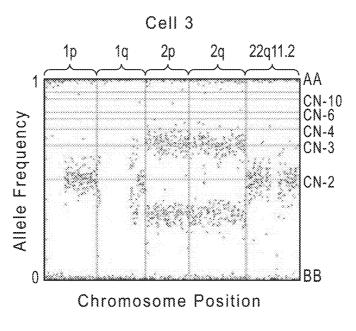
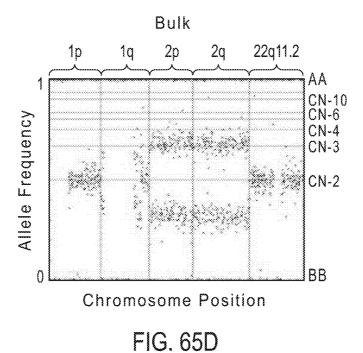
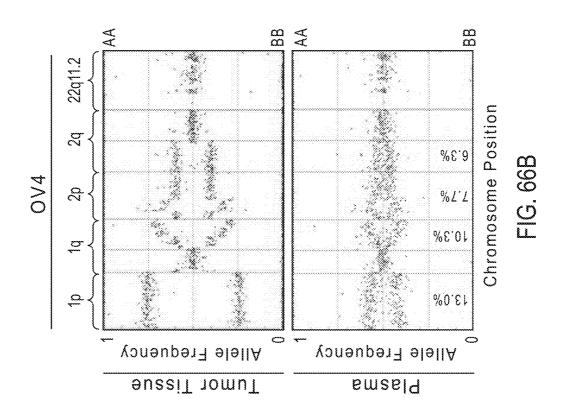
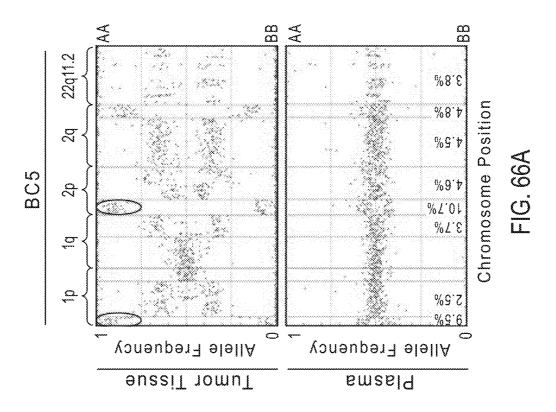


FIG. 65C



U.S. Patent Dec. 20, 2022 Sheet 93 of 105 US 11,530,454 B2





U.S. Patent Dec. 20, 2022 Sheet 94 of 105 US 11,530,454 B2

	1p	1q	2p	2q	22q11.2
OV1					
OV2					
OV3					
OV4					
OV5					

CNV in tissue and matched plasma		
CNV in tumor tissue only		
No CNV detected		

FIG. 66C

U.S. Patent Dec. 20, 2022 Sheet 95 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
9:163985-163985_T>C	9:5050714-5050714_T>C	17:7577587-7577587_A>T
1:17275337-17275337_C>T	9:21971111-21971111_C>T	17:7577590-7577590_A>T
1:23689289-23689289_T>G	9:32632175-32632175_C>T	17:7577592-7577592_C>A
1:26774085-26774085_G>A	9:108097968-108097968_G>A	17:7577598-7577598_A>G
1:36226206-36226206_A>G	9:130931780-130931780_C>T	17:7577599-7577599_G>C
1:43032078-43032078_C>T	9:141071078-141071078_A>G	17:7577604-7577604_G>T
1:43912067-43912067_G>A	10:37486216-37486216_C>G	17:7578177-7578177_G>C
1:64515405-64515405_C>T	10:46254776-46254776_A>C	17:7578183-7578183_G>T
1:99771528-99771528_G>A	10:47207813-47207813_A>G	17:7578188-7578188_G>T
1:110173662-110173662_G>A	10:50678650-50678650_A>T	17:7578190-7578190_A>G
1:115256528-115256528_A>C	10:51853633-51853633_C>T	17:7578191-7578191_T>C
1:115256529-115256529_A>G	10:72358324-72358324_C>T	17:7578196-7578196_T>A
1:120458147-120458147_C>T	10:81058322-81058322_C>G	17:7578202-7578202_T>A
1:145323656-145323656_A>T	10:89624275-89624275_C>T	17:7578203-7578203_G>T
1:148343792-148343792_G>T	10:89624305-89624305_T>G	17:7578204-7578204_T>G
1:151149263-151149263_A>G	10:89692790-89692790_G>A	17:7578206-7578206_A>G
1:152186837-152186837_G>A	10:89692794-89692794_A>T	17:7578207-7578207_T>C
1:152189016-152189016_C>T	10:89692835-89692835_G>A	17:7578208-7578208_A>G
1:152328936-152328936_C>G	10:89692839-89692839_T>G	17:7578212-7578212_C>T
1:157805906-157805906_G>A	10:89692883-89692883_C>G	17:7578215-7578215_T>A
1:158609792-158609792_A>T	10:89692893-89692893_C>A	17:7578216-7578216_T>C
1:186276240-186276240_T>C	10:89692900-89692900_G>C	17:7578224-7578224_A>T
1:198675866-198675866_A>C	10:89692905-89692905_G>A	17:7578225-7578225_C>A
1:202699028-202699028_A>G	10:89692923-89692923_G>A	17:7578235-7578235_A>G
1:225306960-225306960_T>G	10:89692980-89692980_A>G	17:7578236-7578236_T>G
1:227288919-227288919_C>T	10:89692998-89692998_G>T	17:7578239-7578239_G>T

FIG. 67A

U.S. Patent Dec. 20, 2022 Sheet 96 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
1:228560700-228560700_T>C	10:123258034-123258034_T>A	17:7578244-7578244_G>C
1:235884036-235884036_A>G	10:127548395-127548395_C>G	17:7578245-7578245_C>A
1:247614573-247614573_C>T	11:534288-534288_G>A	17:7578253-7578253_G>T
1:248402566-248402566_T>A	11:1093430-1093430_C>A	17:7578254-7578254_G>A
2:32820108-32820108_A>C	11:1643049-1643049_T>G	17:7578257-7578257_G>T
2:61719472-61719472_G>A	11:1651157-1651157_A>G	17:7578259-7578259_T>G
2:107049425-107049425_T>G	11:48171648-48171648_A>T	17:7578260-7578260_G>T
2:107460402-107460402_G>A	11:49854989-49854989_G>A	17:7578262-7578262_G>C
2:118715997-118715997_C>G	11:55541605-55541605_G>A	17:7578263-7578263_C>T
2:119752091-119752091_G>A	11:55861593-55861593_G>A	17:7578265-7578265_T>C
2:129075877-129075877_C>A	11:60666746-60666746_A>C	17:7578266-7578266_A>T
2:165986535-165986535_A>G	11:64544046-64544046_A>G	17:7578268-7578268_T>C
2:178098803-178098803_G>A	11:66335548-66335548_T>C	17:7578269-7578269_C>T
2:179434772-179434772_C>T	11:71932638-71932638_C>T	17:7578271-7578271_A>G
2:179635206-179635206_C>G	11:94204875-94204875_T>G	17:7578272-7578272_C>T
2:197641324-197641324_C>G	11:103182692-103182692_G>A	17:7578275-7578275_C>T
2:198266834-198266834_A>G	11:108117798-108117798_C>T	17:7578280-7578280_C>T
2:198363501-198363501_G>A	11:108175462-108175462_G>A	17:7578281-7578281_C>T
2:209113113-209113113_C>T	11:108200961-108200961_G>A	17:7578284-7578284_G>T
2:238672328-238672328_C>T	12:4627253-4627253_G>A	17:7578371-7578371_G>A
3:1269641-1269641_A>G	12:9453702-9453702_C>T	17:7578374-7578374_G>C
3:19924193-19924193_G>A	12:9574020-9574020_A>G	17:7578375-7578375_C>G
3:25832827-25832827_T>A	12:9583286-9583286_T>C	17:7578380-7578380_G>A
3:38104250-38104250_C>G	12:12630318-12630318_C>T	17:7578382-7578382_C>T
3:41266100-41266100_T>A	12:21953991-21953991_G>A	17:7578388-7578388_G>C
3:41266101-41266101_C>G	12:22677465-22677465_G>A	17:7578389-7578389_C>T

FIG. 67B

U.S. Patent Dec. 20, 2022 Sheet 97 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
3:41266104-41266104_G>T	12:25380275-25380275_A>T	17:7578392-7578392_G>C
3:41266113-41266113_C>A	12:25380276-25380276_A>T	17:7578393-7578393_T>A
3:41266124-41266124_A>G	12:25380277-25380277_C>A	17:7578394-7578394_A>G
3:44612650-44612650_G>T	12:25398281-25398281_G>A	17:7578395-7578395_C>A
3:58191274-58191274_G>C	12:25398282-25398282_G>T	17:7578400-7578400_C>G
3:108788596-108788596_C>T	12:25398284-25398284_G>A	17:7578402-7578402_C>A
3:138022433-138022433_C>T	12:25398285-25398285_G>C	17:7578403-7578403_G>A
3:149245633-149245633_T>G	12:30888067-30888067_G>A	17:7578404-7578404_T>A
3:172533490-172533490_A>C	12:31231425-31231425_T>G	17:7578406-7578406_G>A
3:178916854-178916854_G>A	12:31242869-31242869_G>A	17:7578407-7578407_C>A
3:178916876-178916876_G>A	12:50745677-50745677_A>T	17:7578409-7578409_G>A
3:178916936-178916936_G>A	12:50746243-50746243_A>C	17:7578410-7578410_A>T
3:178916941-178916941_G>A	12:52385715-52385715_C>T	17:7578413-7578413_G>T
3:178916944-178916944_A>G	12:56482341-56482341_G>T	17:7578415-7578415_T>A
3:178916946-178916946_G>C	12:91502039-91502039_G>A	17:7578416-7578416_G>A
3:178917478-178917478_G>A	12:101745882-101745882_A>T	17:7578419-7578419_G>A
3:178921548-178921548_G>A	13:28919595-28919595_G>A	17:7578420-7578420_G>A
3:178921552-178921552_A>T	13:32968854-32968854_C>A	17:7578423-7578423_G>A
3:178921553-178921553_T>A	13:36700097-36700097_C>T	17:7578428-7578428_C>T
3:178927974-178927974_G>A	13:48953760-48953760_C>T	17:7578431-7578431_C>T
3:178927980-178927980_T>C	13:49033955-49033955_A>T	17:7578433-7578433_C>G
3:178928079-178928079_G>A	14:20215706-20215706_G>T	17:7578437-7578437_C>T
3:178936050-178936050_T>C	14:21873400-21873400_G>A	17:7578440-7578440_A>G
3:178936074-178936074_C>G	14:44973867-44973867_C>T	17:7578442-7578442_A>C
3:178936082-178936082_G>A	14:63784405-63784405_C>T	17:7578443-7578443_T>A
3:178936083-178936083_A>T	14:102323093-102323093_T>C	17:7578448-7578448_C>A

FIG. 67C

U.S. Patent Dec. 20, 2022 Sheet 98 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
3:178936091-178936091_G>A	14:105246551~105246551_G>A	17:7578449-7578449_G>A
3:178936092-178936092_A>C	15:22743235-22743235_A>G	17:7578452-7578452_A>G
3:178936093-178936093_G>C	15:33954810-33954810_G>A	17:7578454-7578454_C>T
3:178936094-178936094_C>A	15:66774131-66774131_G>A	17:7578455-7578455_G>C
3:178936095-178936095_A>G	15:75047325-75047325_G>A	17:7578457-7578457_G>A
3:178936097-178936097_G>A	15:83014132-83014132_G>C	17:7578460-7578460_T>G
3:178937019-178937019_A>G	15:84909434-84909434_G>A	17:7578461-7578461_G>T
3:178938860-178938860_A>C	15:102516424-102516424_G>T	17:7578463-7578463_G>C
3:178938934-178938934_G>A	16:1004605-1004605_T>C	17:7578464-7578464_C>G
3:178947827-178947827_G>T	16:2049640-2049640_T>C	17:7578466-7578466_C>T
3:178951957-178951957_G>A	16:3820773-3820773_C>T	17:7578467-7578467_A>C
3:178951964-178951964_G>C	16:4432029-4432029_A>C	17:7578469-7578469_G>A
3:178952004-178952004_C>T	16:28507445-28507445_G>C	17:7578471-7578471_C>T
3:178952013-178952013_G>T	16:29110458-29110458_T>C	17:7578472-7578472_C>T
3:178952018-178952018_A>K	16:56782199-56782199_G>A	17:7578475-7578475_C>T
3:178952019-178952019_C>T	16:68772218-68772218_C>T	17:7578476-7578476_C>T
3:178952020-178952020_C>T	16:68844139-68844139_G>A	17:7578478-7578478_C>A
3:178952030-178952030_G>C	17:7573982-7573982_G>T	17:7578479-7578479_C>T
3:178952072-178952072_A>G	17:7574002-7574002_G>C	17:7578490-7578490_T>A
3:178952074-178952074_G>C	17:7574003-7574003_C>T	17:7578492-7578492_G>A
3:178952075-178952075_A>T	17:7574012-7574012_G>T	17:7578493-7578493_G>A
3:178952077-178952077_T>A	17:7574017-7574017_G>T	17:7578496-7578496_T>A
3:178952081-178952081_G>A	17:7574018-7574018_C>T	17:7578498-7578498_G>T
3:178952082-178952082_C>T	17:7576855-7576855_C>T	17:7578499-7578499_A>C
3:178952084-178952084_C>T	17:7576857-7576857_T>G	17:7578500-7578500_C>T
3:178952085-178952085_A>G	17:7576883-7576883_A>G	17:7578502-7578502_T>C

FIG. 67D

U.S. Patent Dec. 20, 2022 Sheet 99 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
3:178952088-178952088_A>G	17:7576897-7576897_C>T	17:7578506-7578506_C>A
3:178952090-178952090_G>C	17:7577022-7577022_C>T	17:7578507-7578507_C>A
3:178952091-178952091_G>C	17:7577036-7577036_C>A	17:7578508-7578508_G>A
3:178952100-178952100_C>A	17:7577046-7577046_G>T	17:7578509-7578509_T>C
3:192516720-192516720_C>G	17:7577052-7577052_C>A	17:7578513-7578513_G>A
3:195452649-195452649_C>T	17:7577058-7577058_G>T	17:7578514-7578514_A>G
3:195505836-195505836_C>G	17:7577069-7577069_G>A	17:7578517-7578517_C>T
3:195506597-195506597_C>T	17:7577079-7577079_G>T	17:7578518-7578518_G>A
3:195506940-195506940_C>G	17:7577081-7577081_A>C	17:7578524-7578524_C>T
3:195511945-195511945_C>T	17:7577082-7577082_G>C	17:7578525-7578525_C>G
4:11400898-11400898_G>A	17:7577085-7577085_G>A	17:7578526-7578526_G>A
4:15443812-15443812_C>A	17:7577088-7577088_A>C	17:7578527-7578527_T>G
4:40356408-40356408_C>A	17:7577090-7577090_G>C	17:7578528-7578528_T>G
4:46930361-46930361_A>C	17:7577091-7577091_C>T	17:7578529-7578529_T>G
4:48597671-48597671_G>A	17:7577093-7577093_G>A	17:7578530-7578530_T>C
4:55593613-55593613_T>A	17:7577094-7577094_C>T	17:7578532-7578532_T>A
4:55593661-55593661_T>C	17:7577095-7577095_C>A	17:7578534-7578534_G>T
4:55594212-55594212_T>C	17:7577096-7577096_A>G	17:7578535-7578535_A>G
4:55594221-55594221_A>G	17:7577097-7577097_G>A	17:7578536-7578536_A>G
4:55599320-55599320_G>C	17:7577099-7577099_G>A	17:7578541-7578541_T>G
4:71469002-71469002_C>G	17:7577100-7577100_A>T	17:7578542-7578542_C>T
4:83778206-83778206_C>T	17:7577102-7577102_G>A	17:7578548-7578548_C>T
4:88537243-88537243_C>A	17:7577105-7577105_C>G	17:7578550-7578550_C>G
4:88537249-88537249_T>C	17:7577106-7577106_C>A	17:7578551-7578551_T>A
4:107845202-107845202_G>A	17:7577107-7577107_T>G	17:7578553-7578553_A>G
4:144336805-144336805_A>G	17:7577108-7577108_G>T	17:7578554-7578554_T>G

FIG. 67E

U.S. Patent Dec. 20, 2022 Sheet 100 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
4:153245446-153245446_C>T	17:7577111-7577111_C>T	17:7579312-7579312_G>C
4:153247168-153247168_A>G	17:7577112-7577112_G>C	17:7579313-7579313_C>G
4:153247288-153247288_G>H	17:7577113-7577113_T>G	17:7579329-7579329_A>G
4:153247366-153247366_G>T	17:7577114-7577114_G>T	17:7579349-7579349_T>G
4:153249456-153249456_G>T	17:7577115-7577115_T>G	17:7579355-7579355_T>C
4:164507069-164507069_C>T	17:7577117-7577117_T>C	17:7579358-7579358_G>C
4:170321765-170321765_C>T	17:7577118-7577118_G>T	17:7579366-7579366_C>G
4:175896768-175896768_C>T	17:7577120-7577120_G>A	17:7579377-7579377_C>T
5:13762932-13762932_G>A	17:7577121-7577121_C>T	17:7579378-7579378_C>A
5:26906161-26906161_G>A	17:7577123-7577123_T>C	17:7579414-7579414_G>A
5:67591128-67591128_G>C	17:7577124-7577124_G>T	17:7579503-7579503_G>T
5:112173917-112173917_C>T	17:7577127-7577127_G>T	17:7579521-7579521_G>A
5:112175118-112175118_C>T	17:7577129-7577129_T>G	17:7579547-7579547_C>T
5:112175219-112175219_A>T	17:7577130-7577130_T>C	17:7579882-7579882_G>C
5:112175681-112175681_G>T	17:7577138-7577138_G>A	17:11554600-11554600_G>T
5:112175786-112175786_G>T	17:7577141-7577141_G>A	17:11998898-11998898_C>T
5:112176020-112176020_G>T	17:7577142-7577142_G>T	17:12011144-12011144_C>T
5:126791225-126791225_C>T	17:7577144-7577144_T>C	17:37868208-37868208_C>T
5:129030517-129030517_C>T	17:7577148-7577148_C>G	17:37880220-37880220_T>C
5:131931452-131931452_A>T	17:7577501-7577501_C>T	17:37880261-37880261_G>T
5:135394826-135394826_G>A	17:7577505-7577505_A>T	17:37881000-37881000_G>C
5:140683389-140683389_C>T	17:7577506-7577506_G>T	17:37881332-37881332_G>A
5:170239186-170239186_A>G	17:7577508-7577508_A>G	17:39240627-39240627_T>C
5:180335598-180335598_T>G	17:7577509-7577509_G>A	17:39274087-39274087_C>G
6:10756728-10756728_C>T	17:7577511-7577511_T>A	17:39274157-39274157_C>T
6:26032069-26032069_G>A	17:7577517-7577517_T>C	17:39305800-39305800_A>T

FIG. 67F

U.S. Patent Dec. 20, 2022 Sheet 101 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
6:28497279-28497279_G>A	17:7577518-7577518_A>T	17:39595484-39595484_C>T
6:29910622-29910622_C>T	17:7577520-7577520_T>C	17:39673185-39673185_G>A
6:93956553-93956553_C>G	17:7577521-7577521_A>G	17:41245274-41245274_G>C
6:112452259-112452259_G>A	17:7577524-7577524_A>C	17:45219311-45219311_A>G
6:130497111-130497111_C>T	17:7577526-7577526_T>C	17:48777241-48777241_T>A
6:150001395-150001395_C>T	17:7577529-7577529_T>A	17:58288421-58288421_G>A
6:170871013-170871013_A>G	17:7577532-7577532_C>A	17:61958402-61958402_C>T
7:2577781-2577781_A>G	17:7577533-7577533_C>G	18:29470816-29470816_G>A
7:6426892-6426892_C>T	17:7577534-7577534_G>A	18:45368211-45368211_C>G
7:27832791-27832791_A>G	17:7577535-7577535_G>C	18:48575671-48575671_C>G
7:43591930-43591930_G>A	17:7577536-7577536_A>T	18:48604754-48604754_G>T
7:55259482-55259482_C>A	17:7577538-7577538_G>A	19:1223125-1223125_C>G
7:55259515-55259515_T>G	17:7577539-7577539_C>T	19:22836805-22836805_G>A
7:77379331-77379331_T>G	17:7577541-7577541_A>T	19:40383905-40383905_C>T
7:100643088-100643088_A>G	17:7577543-7577543_G>A	19:49926533-49926533_C>G
7:100677279-100677279_G>C	17:7577544-7577544_T>G	19:51274851-51274851_A>C
7:138268672-138268672_C>T	17:7577545-7577545_A>G	19:56369444-56369444_G>A
7:140453136-140453136_T>A	17:7577547-7577547_G>T	20:3209652-3209652_T>G
7:140453137-140453137_G>A	17:7577548-7577548_G>A	20:6751034-6751034_A>G
7:140481411-140481411_G>T	17:7577549-7577549_C>G	20:29623210-29623210_C>T
7:140481412-140481412_G>C	17:7577550-7577550_G>T	20:29623224-29623224_C>A
7:140481417-140481417_G>T	17:7577551-7577551_G>C	20:29628236-29628236_G>C
7:140500165-140500165_T>C	17:7577552-7577552_G>C	20:29628263-29628263_A>G
7:151970859-151970859_G>A	17:7577554-7577554_A>T	20:29632643-29632643_T>C
7:151970951-151970951_G>A	17:7577555-7577555_C>A	20:31671497-31671497_C>T
8:2796250-2796250_C>T	17:7577556-7577556_G>A	20:34242060~34242060_G>A

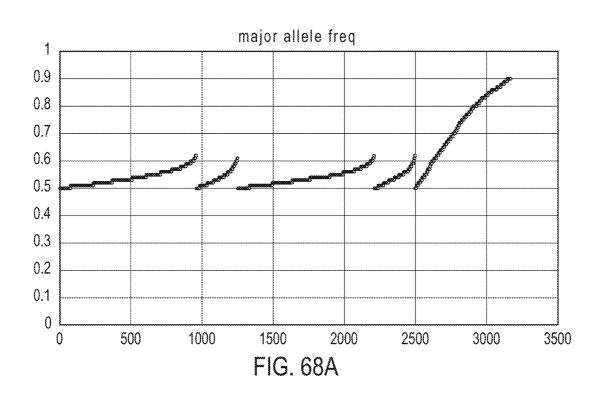
FIG. 67G

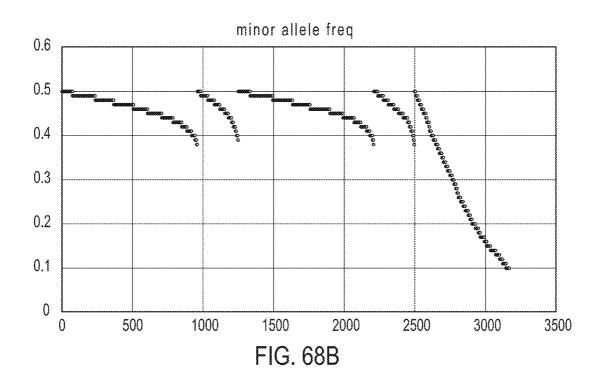
U.S. Patent Dec. 20, 2022 Sheet 102 of 105 US 11,530,454 B2

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
8:2855562-2855562_A>G	17:7577557-7577557_T>A	20:57484420-57484420_C>T
8:3009020-3009020_C>G	17:7577559-7577559_C>A	20:57484421-57484421_G>A
8:3205603-3205603_G>A	17:7577565-7577565_A>C	21:10951281-10951281_G>A
8:3245034-3245034_G>T	17:7577566-7577566_A>G	21:32127654-32127654_G>A
8:10467948-10467948_C>T	17:7577568-7577568_G>A	22:16449539-16449539_T>C
8:77768353-77768353_C>T	17:7577569-7577569_T>A	22:22899234-22899234_T>C
8:89086854-89086854_G>A	17:7577570-7577570_G>A	22:29092948-29092948_C>T
8:101083610-101083610_T>C	17:7577571-7577571_T>A	22:30398972-30398972_G>C
8:109796543-109796543_G>A	17:7577574-7577574_A>G	23:15821891-15821891_C>G
8:116426722-116426722_G>T	17:7577575-7577575_T>A	23:66765161-66765161_A>T
8:121344960-121344960_G>A	17:7577577-7577577_A>G	23:78011285-78011285_G>A
8:144940290-144940290_G>C	17:7577580-7577580_A>G	23:105153675-105153675_C>A
8:144940331-144940331_G>A	17:7577581~7577581_T>A	17:7577586-7577586_T>G

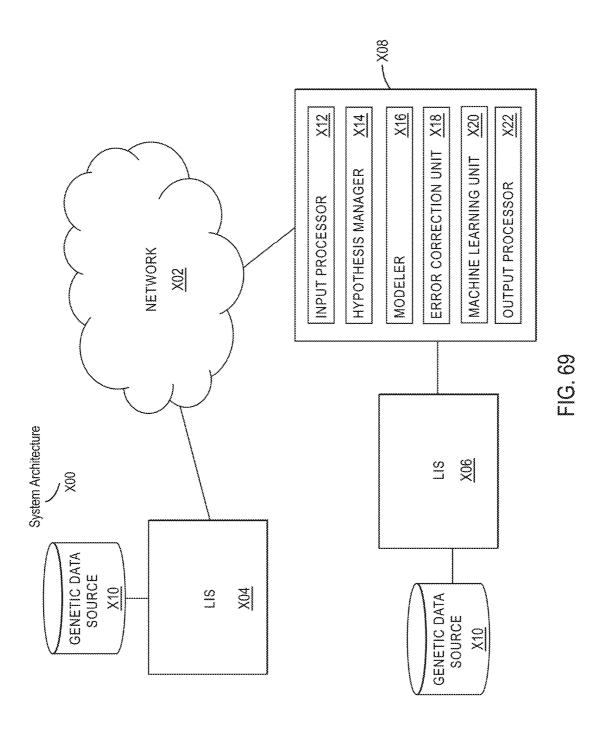
FIG. 67H

U.S. Patent Dec. 20, 2022 Sheet 103 of 105 US 11,530,454 B2

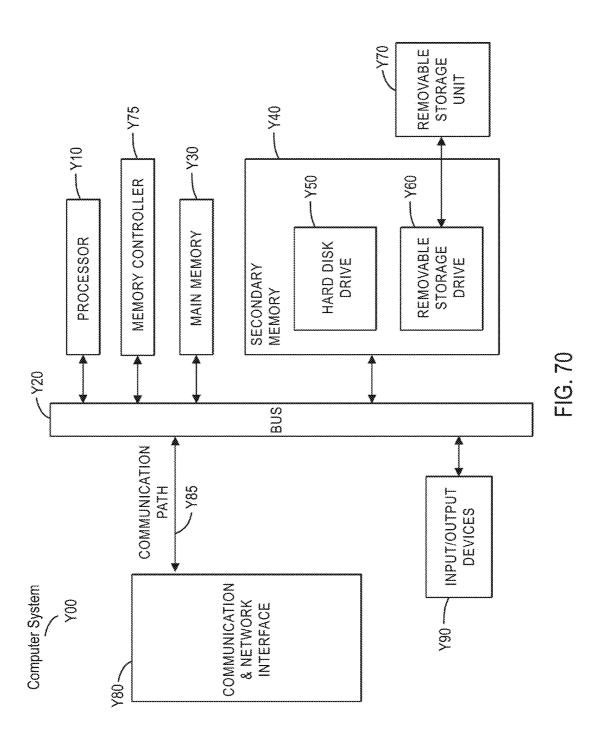




U.S. Patent Dec. 20, 2022 Sheet 104 of 105 US 11,530,454 B2



U.S. Patent Dec. 20, 2022 Sheet 105 of 105 US 11,530,454 B2



## US 11,530,454 B2

#### 1

# DETECTING MUTATIONS AND PLOIDY IN CHROMOSOMAL SEGMENTS

# CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. Utility application Ser. No. 17/692,469, filed Mar. 11, 2022. U.S. Utility application Ser. No. 17/692,469 is a continuation of U.S. Utility application Ser. No. 15/898,145, filed Feb. 15, 2018 (now U.S. Pat. No. 11,319,595). U.S. Utility application Ser. No. 15/898,145 is a continuation of U.S. Utility application Ser. No. 14/692,703, filed Apr. 21, 2015 (now U.S. Pat. No. 10,179,937), which claims the benefit of and priority to U.S.  $_{15}$ Provisional Application Ser. No. 61/982,245, filed Apr. 21, 2014; U.S. Provisional Application Ser. No. 61/987,407, filed May 1, 2014; U.S. Provisional Application Ser. No. 62/066,514, filed Oct. 21, 2014; U.S. Provisional Application Ser. No. 62/146,188, filed Apr. 10, 2015; U.S. Provi- 20 sional Application Ser. No. 62/147,377, filed Apr. 14, 2015; U.S. Provisional Application Ser. No. 62/148,173, filed Apr. 15, 2015, the entirety of these applications are hereby incorporated herein by reference for the teachings therein.

#### FIELD OF THE INVENTION

The present invention generally relates to methods and systems for detecting ploidy of a chromosome segment, and methods and systems for detecting a single nucleotide <sup>30</sup> variant.

#### BACKGROUND OF THE INVENTION

Copy number variation (CNV) has been identified as a 35 major cause of structural variation in the genome, involving both duplications and deletions of sequences that typically range in length from 1,000 base pairs (1 kb) to 20 megabases (mb). Deletions and duplications of chromosome segments or entire chromosomes are associated with a variety of 40 conditions, such as susceptibility or resistance to disease.

CNVs are often assigned to one of two main categories, based on the length of the affected sequence. The first category includes copy number polymorphisms (CNPs), which are common in the general population, occurring with 45 an overall frequency of greater than 1%. CNPs are typically small (most are less than 10 kilobases in length), and they are often enriched for genes that encode proteins important in drug detoxification and immunity. A subset of these CNPs is highly variable with respect to copy number. As a result, 50 different human chromosomes can have a wide range of copy numbers (e.g., 2, 3, 4, 5, etc.) for a particular set of genes. CNPs associated with immune response genes have recently been associated with susceptibility to complex genetic diseases, including psoriasis, Crohn's disease, and 55 glomerulonephritis.

The second class of CNVs includes relatively rare variants that are much longer than CNPs, ranging in size from hundreds of thousands of base pairs to over 1 million base pairs in length. In some cases, these CNVs may have arisen 60 during production of the sperm or egg that gave rise to a particular individual, or they may have been passed down for only a few generations within a family. These large and rare structural variants have been observed disproportionately in subjects with mental retardation, developmental 65 delay, schizophrenia, and autism. Their appearance in such subjects has led to speculation that large and rare CNVs may

2

be more important in neurocognitive diseases than other forms of inherited mutations, including single nucleotide substitutions

Gene copy number can be altered in cancer cells. For instance, duplication of Chr1p is common in breast cancer, and the EGFR copy number can be higher than normal in non-small cell lung cancer. Cancer is one of the leading causes of death; thus, early diagnosis and treatment of cancer is important, since it can improve the patient's outcome (such as by increasing the probability of remission and the duration of remission). Early diagnosis can also allow the patient to undergo fewer or less drastic treatment alternatives. Many of the current treatments that destroy cancerous cells also affect normal cells, resulting in a variety of possible side-effects, such as nausea, vomiting, low blood cell counts, increased risk of infection, hair loss, and ulcers in mucous membranes. Thus, early detection of cancer is desirable since it can reduce the amount and/or number of treatments (such as chemotherapeutic agents or radiation) needed to eliminate the cancer.

Copy number variation has also been associated with severe mental and physical handicaps, and idiopathic learning disability. Non-invasive prenatal testing (NIPT) using cell-free DNA (cfDNA) can be used to detect abnormalities, such as fetal trisomies 13, 18, and 21, triploidy, and sex chromosome aneuploidies. Subchromosomal microdeletions, which can also result in severe mental and physical handicaps, are more challenging to detect due to their smaller size. Eight of the microdeletion syndromes have an aggregate incidence of more than 1 in 1000, making them nearly as common as fetal autosomal trisomies.

In addition, a higher copy number of CCL3L1 has been associated with lower susceptibility to HIV infection, and a low copy number of FCGR3B (the CD16 cell surface immunoglobulin receptor) can increase susceptibility to systemic lupus erythematosus and similar inflammatory autoimmune disorders.

Thus, improved methods are needed to detect deletions and duplications of chromosome segments or entire chromosomes. Preferably, these methods can be used to more accurately diagnose disease or an increased risk of disease, such as cancer or CNVs in a gestating fetus.

#### SUMMARY OF THE INVENTION

In illustrative embodiments, provided herein is a method for determining ploidy of a chromosomal segment in a sample of an individual. The method includes the following steps:

- a. receiving allele frequency data comprising the amount of each allele present in the sample at each loci in a set of polymorphic loci on the chromosomal segment;
- generating phased allelic information for the set of polymorphic loci by estimating the phase of the allele frequency data;
- c. generating individual probabilities of allele frequencies for the polymorphic loci for different ploidy states using the allele frequency data;
- d. generating joint probabilities for the set of polymorphic loci using the individual probabilities and the phased allelic information; and
- e. selecting, based on the joint probabilities, a best fit model indicative of chromosomal ploidy, thereby determining ploidy of the chromosomal segment.

In one illustrative embodiment of the method for determining ploidy, the data is generated using nucleic acid sequence data, especially high throughput nucleic acid

## US 11,530,454 B2

sequence data. In certain illustrative examples of the method for determining ploidy, the allele frequency data is corrected for errors before it is used to generate individual probabilities. In specific illustrative embodiments, the errors that are corrected include allele amplification efficiency bias. In 5 other embodiments, the errors that are corrected include ambient contamination and genotype contamination. In some embodiments, errors that are corrected include allele amplification bias, ambient contamination and genotype

In certain embodiments of the method for determining ploidy, the individual probabilities are generated using a set of models of both different ploidy states and allelic imbalance fractions for the set of polymorphic loci. In these embodiments, and other embodiments, the joint probabili- 15 ties are generated by considering the linkage between polymorphic loci on the chromosome segment.

Accordingly, in one illustrative embodiment that combines some of these embodiments, provided herein is a method for detecting chromosomal ploidy in a sample of an 20 individual, that includes the following steps:

- a. receiving nucleic acid sequence data for alleles at a set of polymorphic loci on a chromosome segment in the
- b. detecting allele frequencies at the set of loci using the 25 nucleic acid sequence data;
- c. correcting for allele amplification efficiency bias in the detected allele frequencies to generate corrected allele frequencies for the set of polymorphic loci;
- d. generating phased allelic information for the set of 30 polymorphic loci by estimating the phase of the nucleic acid sequence data;
- e. generating individual probabilities of allele frequencies for the polymorphic loci for different ploidy states by comparing the corrected allele frequencies to a set of 35 models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci;
- f. generating joint probabilities for the set of polymorphic loci by combining the individual probabilities considering the linkage between polymorphic loci on the 40 chromosome segment; and
- g. selecting, based on the joint probabilities, the best fit model indicative of chromosomal aneuploidy.

In another aspect, provided herein is a system for detecting chromosomal ploidy in a sample of an individual, the 45 method for determining whether circulating tumor nucleic system comprising:

- a. an input processor configured to receive allelic frequency data comprising the amount of each allele present in the sample at each loci in a set of polymorphic loci on the chromosomal segment;
- b. a modeler configured to:
  - i. generate phased allelic information for the set of polymorphic loci by estimating the phase of the allele frequency data; and
  - ii. generate individual probabilities of allele frequen- 55 cies for the polymorphic loci for different ploidy states using the allele frequency data; and
  - iii. generate joint probabilities for the set of polymorphic loci using the individual probabilities and the phased allelic information; and
- c. a hypothesis manager configured to select, based on the joint probabilities, a best fit model indicative of chromosomal ploidy, thereby determining ploidy of the chromosomal segment.

In certain embodiments of this system embodiment, the 65 allele frequency data is data generated by a nucleic acid sequencing system. In certain embodiments, the system

further comprises an error correction unit configured to correct for errors in the allele frequency data, wherein the corrected allele frequency data is used by the modeler for to generate individual probabilities. In certain embodiments the error correction unit corrects for allele amplification efficiency bias. In certain embodiments, the modeler generates the individual probabilities using a set of models of both different ploidy states and allelic imbalance fractions for the set of polymorphic loci. The modeler, in certain exemplary embodiments generates the joint probabilities by considering the linkage between polymorphic loci on the chromosome segment.

In one illustrative embodiment, provided herein is a system for detecting chromosomal ploidy in a sample of an individual, that includes the following:

- a. an input processor configured to receive nucleic acid sequence data for alleles at a set of polymorphic loci on a chromosome segment in the individual and detect allele frequencies at the set of loci using the nucleic acid sequence data;
- b. an error correction unit configured to correct for errors in the detected allele frequencies and generate corrected allele frequencies for the set of polymorphic loci:
- c. a modeler configured to:
  - i. generate phased allelic information for the set of polymorphic loci by estimating the phase of the nucleic acid sequence data;
  - ii. generate individual probabilities of allele frequencies for the polymorphic loci for different ploidy states by comparing the phased allelic information to a set of models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci;
  - iii. generate joint probabilities for the set of polymorphic loci by combining the individual probabilities considering the relative distance between polymorphic loci on the chromosome segment; and
- d. a hypothesis manager configured to select, based on the joint probabilities, a best fit model indicative of chromosomal aneuploidy.

In certain aspects, the present invention provides a acids are present in a sample in an individual, comprising

- a. analyzing the sample to determine a ploidy at a set of polymorphic loci on a chromosome segment in the individual: and
- b. determining the level of allelic imbalance present at the polymorphic loci based on the ploidy determination, wherein an allelic imbalance equal to or greater than 0.4%, 0.45%, or 0.5% is indicative of the presence of circulating tumor nucleic acids in the sample.

In certain embodiments the method for determining whether circulating tumor nucleic acids are present, further comprises detecting a single nucleotide variant at a single nucleotide variance site in a set of single nucleotide variance locations, wherein detecting either an allelic imbalance equal to or greater than 45% or detecting the single nucleotide variant, or both, is indicative of the presence of circulating tumor nucleic acids in the sample.

In certain embodiments analyzing step in the method for determining whether circulating tumor nucleic acids are present, includes analyzing a set of chromosome segments known to exhibit aneuploidy in cancer. In certain embodiments analyzing step in the method for determining whether

## US 11,530,454 B2

4

circulating tumor nucleic acids are present, includes analyzing between 1,000 and 50,000 or between 100 and 1000, polymorphic loci for ploidy.

In certain aspects, provided herein are methods for detecting single nucleotide variants in a sample. Accordingly, provided herein is a method for determining whether a single nucleotide variant is present at a set of genomic positions in a sample from an individual, the method comprising:

- a. for each genomic position, generating an estimate of 10 efficiency and a per cycle error rate for an amplicon spanning that genomic position, using a training data set:
- receiving observed nucleotide identity information for each genomic position in the sample;
- c. determining a set of probabilities of single nucleotide variant percentage resulting from one or more real mutations at each genomic position, by comparing the observed nucleotide identity information at each genomic position to a model of different variant percentages using the estimated amplification efficiency and the per cycle error rate for each genomic position independently; and
- d. determining the most-likely real variant percentage and confidence from the set of probabilities for each 25 genomic position.

In illustrative embodiments of the method for determining whether a single nucleotide variant is present, the estimate of efficiency and the per cycle error rate is generated for a set of amplicons that span the genomic position. For 30 example, 2, 3, 4, 5, 10, 15, 20, 25, 50, 100 or more amplicons can be included that span the genomic position. In certain embodiments of this method for detecting one or more SNVs the limit of detection is 0.015%, 0.017%, or 0.02%.

In illustrative embodiments of the method for determining whether a single nucleotide variant is present, the observed nucleotide identity information comprises an observed number of total reads for each genomic position and an observed number of variant allele reads for each genomic position.

In illustrative embodiments of the method for determining 40 whether a single nucleotide variant is present, the sample is a plasma sample and the single nucleotide variant is present in circulating tumor DNA of the sample.

In another embodiment, provided herein is a method for detecting one or more single nucleotide variants in a test 45 sample from an individual. The method according to this embodiment, includes the following steps:

- a. determining a median variant allele frequency for a plurality of control samples from each of a plurality of normal individuals, for each single nucleotide variant 50 position in a set of single nucleotide variance positions based on results generated in a sequencing run, to identify selected single nucleotide variant positions having variant median allele frequencies in normal samples below a threshold value and to determine 55 background error for each of the single nucleotide variant positions after removing outlier samples for each of the single nucleotide variant positions;
- b. determining an observed depth of read weighted mean and variance for the selected single nucleotide variant 60 positions for the test sample based on data generated in the sequencing run for the test sample; and
- c. identifying using a computer, one or more single nucleotide variant positions with a statistically significant depth of read weighted mean compared to the 65 background error for that position, thereby detecting the one or more single nucleotide variants.

6

In certain embodiments of this method for detecting one or more SNVs the sample is a plasma sample, the control samples are plasma samples, and the detected one or more single nucleotide variants detected is present in circulating 5 tumor DNA of the sample. In certain embodiments of this method for detecting one or more SNVs the plurality of control samples comprises at least 25 samples. In certain embodiments of this method for detecting one or more SNVs, outliers are removed from the data generated in the 10 high throughput sequencing run to calculate the observed depth of read weighted mean and observed variance are determined. In certain embodiments of this method for detecting one or more SNVs the depth of read for each single nucleotide variant position for the test sample is at least 100 15 reads.

In certain embodiments of this method for detecting one or more SNVs the sequencing run comprises a multiplex amplification reaction performed under limited primer reaction conditions. In certain embodiments of this method for detecting one or more SNVs the limit of detection is 0.015%, 0.017%, or 0.02%.

In one aspect, the invention features a method of determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of one or more cells from an individual. In some embodiments, the method includes obtaining phased genetic data for the first homologous chromosome segment comprising, the identity of the allele present at that locus on the first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, obtaining phased genetic data for the second homologous chromosome segment comprising the identity of the allele present at that locus on the second homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and obtaining measured genetic allelic data comprising the amount of each allele present in a sample of DNA or RNA from one or more cells from the individual, for each of the alleles at each of the loci in the set of polymorphic loci. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment in the genome of one or more cells from the individual, calculating (such as calculating on a computer) a likelihood of one or more of the hypotheses based on the obtained genetic data of the sample and the obtained phased genetic data, and selecting the hypothesis with the greatest likelihood, thereby determining the degree of overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more cells from the individual. In some embodiments, the phased data includes inferred phased data using population based haplotype frequencies and/or measured phased data (e.g., phased data obtained by measuring a sample containing DNA or RNA from the individual or a relative of the individual)

In one aspect, the invention provides a method for determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of one or more cells from an individual. In some embodiments, the method includes obtaining phased genetic data for the first homologous chromosome segment comprising the identity of the allele present at that locus on the first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, obtaining phased genetic data for the second

## US 11,530,454 B2

1

homologous chromosome segment comprising the identity of the allele present at that locus on the second homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and obtaining measured genetic allelic data comprising the amount of each allele present in a sample of DNA or RNA from one or more cells from the individual for each of the alleles at each of the loci in the set of polymorphic loci. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of over- 10 representation of the first homologous chromosome segment; calculating, for each of the hypotheses, expected genetic data for the plurality of loci in the sample from the obtained phased genetic data; calculating (such as calculating on a computer) the data fit between the obtained genetic 15 data of the sample and the expected genetic data for the sample; ranking one or more of the hypotheses according to the data fit; and selecting the hypothesis that is ranked the highest, thereby determining the degree of overrepresentation of the number of copies of the first homologous chro- 20 mosome segment in the genome of one or more cells from the individual.

In one aspect, the invention features a method for determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as com- 25 pared to a second homologous chromosome segment in the genome of one or more cells from an individual. In some embodiments, the method includes obtaining phased genetic data for the first homologous chromosome segment comprising the identity of the allele present at that locus on the 30 first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, obtaining phased genetic data for the second homologous chromosome segment comprising the identity of the allele present at that locus on the second homologous 3 chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and obtaining measured genetic allelic data comprising, for each of the alleles at each of the loci in the set of polymorphic loci, the amount of each allele present in a sample of 40 DNA or RNA from one or more target cells and one or more non-target cells from the individual. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment; calculating 45 (such as calculating on a computer), for each of the hypotheses, expected genetic data for the plurality of loci in the sample from the obtained phased genetic data for one or more possible ratios of DNA or RNA from the one or more target cells to the total DNA or RNA in the sample; calcu- 50 lating (such as calculating on a computer) for each possible ratio of DNA or RNA and for each hypothesis, the data fit between the obtained genetic data of the sample and the expected genetic data for the sample for that possible ratio of DNA or RNA and for that hypothesis; ranking one or 55 more of the hypotheses according to the data fit; and selecting the hypothesis that is ranked the highest, thereby determining the degree of overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more cells from the individual.

In one aspect, the invention features a method for determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of one or more cells from an individual. In some 65 embodiments, the method includes obtaining phased genetic data for the first homologous chromosome segment com-

8

prising the identity of the allele present at that locus on the first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, obtaining phased genetic data for the second homologous chromosome segment comprising the identity of the allele present at that locus on the second homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and obtaining measured genetic allelic data comprising the amount of each allele present in a sample of DNA or RNA from one or more target cells and one or more non-target cells from the individual for each of the alleles at each of the loci in the set of polymorphic loci. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment; calculating (such as calculating on a computer), for each of the hypotheses, expected genetic data for the plurality of loci in the sample from the obtained phased genetic data for one or more possible ratios of DNA or RNA from the one or more target cells to the total DNA or RNA in the sample; calculating (such as calculating on a computer) for each locus in the plurality of loci, each possible ratio of DNA or RNA, and each hypothesis, the likelihood that the hypothesis is correct by comparing the obtained genetic data of the sample for that locus and the expected genetic data for that locus for that possible ratio of DNA or RNA and for that hypothesis; determining the combined probability for each hypothesis by combining the probabilities of that hypothesis for each locus and each possible ratio; and selecting the hypothesis with the greatest combined probability, thereby determining the degree of overrepresentation of the number of copies of the first homologous chromosome segment. In some embodiments, all of the loci are considered at once to calculate the probability of a particular hypothesis, and the hypothesis with the greatest probability is selected.

In one aspect, the invention features a method for determining a number of copies of a chromosome segment of interest in the genome of a fetus. In some embodiments, the method includes obtaining phased genetic data for at least one biological parent of the fetus, wherein the phased genetic data comprises the identity of the allele present for each locus in a set of polymorphic loci on a first homologous chromosome segment and a second homologous chromosome segment in a pair of homologous chromosome segments that comprises the chromosome segment of interest. In some embodiments, the method includes obtaining genetic data at the set of polymorphic loci on the chromosome segment of interest in a mixed sample of DNA or RNA comprising fetal DNA or RNA and maternal DNA or RNA from the mother of the fetus by measuring the quantity of each allele at each locus. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the number of copies of the chromosome segment of interest present in the genome of the fetus. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying, for one or both parents, the number of copies of the first homologous chromosome segment or portion thereof from the parent in the genome of the fetus, the number of copies of the second homologous chromosome segment or portion thereof from the parent in the genome of the fetus, and the total number of copies of the chromosome segment of interest present in the genome of the fetus. In some embodiments, the method includes calculating (such as calculating on a computer), for each of the hypotheses, expected genetic data for the plurality of loci in the mixed sample from the obtained phased genetic data Case: 24-1324 Document: 42-1 Page: 454 Filed: 03/18/2024

## US 11,530,454 B2

sample.

from the parent(s); calculating (such as calculating on a computer) the data fit between the obtained genetic data of the mixed sample and the expected genetic data for the mixed sample; ranking one or more of the hypotheses according to the data fit; and selecting the hypothesis that is ranked the highest, thereby determining the number of copies of the chromosome segment of interest in the genome

In one aspect, the invention features a method for determining a number of copies of a chromosome or chromosome 10 segment of interest in the genome of a fetus. In some embodiments, the method includes obtaining phased genetic data for at least one biological parent of the fetus, wherein the phased genetic data comprises the identity of the allele present for each locus in a set of polymorphic loci on a first 15 homologous chromosome segment and a second homologous chromosome segment in the parent. In some embodiments, the method includes obtaining genetic data at the set of polymorphic loci on the chromosome or chromosome DNA or RNA and maternal DNA or RNA from the mother of the fetus by measuring the quantity of each allele at each locus. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the number of copies of the chromosome or chromosome segment of 25 interest present in the genome of the fetus. In some embodiments, the method includes creating (such as creating on a computer) for each of the hypotheses, a probability distribution of the expected quantity of each allele at each of the plurality of loci in mixed sample from the (i) the obtained 30 phased genetic data from the parent(s) and (ii) optionally the probability of one or more crossovers that may have occurred during the formation of a gamete that contributed a copy of the chromosome or chromosome segment of interest to the fetus; calculating (such as calculating on a 3 computer) a fit, for each of the hypotheses, between (1) the obtained genetic data of the mixed sample and (2) the probability distribution of the expected quantity of each allele at each of the plurality of loci in mixed sample for that hypothesis; ranking one or more of the hypotheses according 40 to the data fit; and selecting the hypothesis that is ranked the highest, thereby determining the number of copies of the chromosome segment of interest in the genome of the fetus.

In some embodiments, the method includes obtaining phased genetic data for the mother of the fetus. In some 45 embodiments, the method includes enumerating a set of one or more hypotheses specifying the number of copies of the first homologous chromosome segment or portion thereof from the mother in the genome of the fetus, the number of copies of the second homologous chromosome segment or 50 portion thereof from the mother in the genome of the fetus, and the total number of copies of the chromosome segment of interest present in the genome of the fetus. In some embodiments, the method includes calculating, for each of the hypotheses, expected genetic data for the plurality of loci 55 in the mixed sample from the obtained phased genetic data from the mother.

In some embodiments, the expected genetic data for each of the hypotheses comprises the identity and an amount of one or more alleles at each locus in the plurality of loci from 60 the maternal DNA or RNA and fetal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating (such as calculating on a computer) expected genetic data by determining a fraction of fetal DNA or RNA and a fraction of maternal DNA or RNA in the mixed 65 sample. In some embodiments, the method includes calculating, for each locus in the plurality of loci, the expected

amount of one or more of the alleles for that locus in the maternal DNA or RNA in the mixed sample using the identity of the allele(s) present at that locus in the obtained phased genetic data of the mother and the fraction of maternal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating (such as calculating on a computer), for each locus in the plurality of loci for each hypothesis, the expected amount of one or more of the alleles for that locus in the fetal DNA or RNA inherited from the mother in the mixed sample using the identity of the allele present at that locus in the first or second homologous chromosome segment from the mother that is specified by the hypothesis to have been inherited by the fetus, the number of copies of the first or second homologous chromosome segment from the mother that is specified by the hypothesis to have been inherited by the fetus, and the fraction of fetal DNA or RNA in the mixed

In some embodiments, the expected genetic data for each segment in a mixed sample of DNA or RNA comprising fetal 20 of the hypotheses comprises the identity and an amount of one or more alleles at each locus in the plurality of loci from the maternal DNA or RNA and fetal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating expected genetic data by determining a fraction of fetal DNA or RNA and a fraction of maternal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating (such as calculating on a computer), for each locus in the plurality of loci, the expected amount of one or more of the alleles for that locus in the maternal DNA or RNA in the mixed sample using the identity of the allele(s) present at that locus in the obtained phased genetic data of the mother and the fraction of maternal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating (such as calculating on a computer), for each locus in the plurality of loci for each hypothesis, the expected amount of one or more of the alleles for that locus in the fetal DNA or RNA inherited from the mother in the mixed sample using the identity of the allele present at that locus in the first or second homologous chromosome segment from the mother that is specified by the hypothesis to have been inherited by the fetus, the number of copies of the first or second homologous chromosome segment from the mother that is specified by the hypothesis to have been inherited by the fetus, the identity of one or more possible alleles at that locus in the first or second homologous chromosome segment from the father that is specified by the hypothesis to have been inherited by the fetus, the number of copies of the first or second homologous chromosome segment from the father that is specified by the hypothesis to have been inherited by the fetus, and the fraction of fetal DNA or RNA in the mixed sample. In some embodiments, population frequencies are used to predict the identity of the alleles in the first or second homologous chromosome segment from the father. In some embodiments, the probability for each of the possible alleles at each locus in the first or second homologous chromosome segment from the father are considered to be the same.

In some embodiments, the method includes obtaining phased genetic data for both the mother and father of the fetus. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the number of copies of the first homologous chromosome segment or portion thereof from the mother in the genome of the fetus, the number of copies of the second homologous chromosome segment or portion thereof from the mother in the genome of the fetus, the number of copies of the first homologous chromosome segment or portion thereof from

## US 11,530,454 B2

1

the father in the genome of the fetus, the number of copies of the second homologous chromosome segment or portion thereof from the father in the genome of the fetus, and the total number of copies of the chromosome segment of interest present in the genome of the fetus. In some embodisments, the method includes calculating (such as calculating on a computer), for each of the hypotheses, expected genetic data for the plurality of loci in the mixed sample from the obtained phased genetic data from the mother and obtained phased genetic data from the father.

In some embodiments, the expected genetic data for each of the hypotheses comprises the identity and an amount of one or more alleles at each locus in the plurality of loci from the maternal DNA or RNA and fetal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating expected genetic data by determining a fraction of fetal DNA or RNA and a fraction of maternal DNA or RNA in the mixed sample. In some embodiments, the method includes calculating (such as calculating on a computer), for each locus in the plurality of loci, the expected 20 amount of one or more of the alleles for that locus in the maternal DNA or RNA in the mixed sample using the identity of the allele(s) present at that locus in the obtained phased genetic data of the mother and the fraction of maternal DNA or RNA in the mixed sample. In some 25 embodiments, the method includes calculating (such as calculating on a computer), for each locus in the plurality of loci for each hypothesis, the expected amount of one or more of the alleles for that locus in the fetal DNA or RNA in the mixed sample using the identity of the allele present at that 30 locus in the first or second homologous chromosome segment from the mother that is specified by the hypothesis to have been inherited by the fetus, the number of copies of the first or second homologous chromosome segment from the mother that is specified by the hypothesis to have been 35 inherited by the fetus, the identity of the allele present at that locus in the first or second homologous chromosome segment from the father that is specified by the hypothesis to have been inherited by the fetus, the number of copies of the first or second homologous chromosome segment from the 40 father that is specified by the hypothesis to have been inherited by the fetus, and the fraction of fetal DNA or RNA in the mixed sample.

In some embodiments, the method includes calculating (such as calculating on a computer), for each of the hypoth- 45 eses, a probability distribution of expected genetic data for the plurality of loci in the mixed sample from the obtained phased genetic data from the parent(s). In some embodiments, the method includes increasing the probability in the probability distribution of an a particular allele being present 50 at a first locus in the mixed sample if that particular allele is present in the first homologous segment in the parent and an allele at a nearby locus in the first homologous segment in the parent is observed in the obtained genetic data of the mixed sample; or decreasing the probability in the probabil- 55 ity distribution of an a particular allele being present at a first locus in the mixed sample if that particular allele is present in the first homologous segment in the parent and an allele at a nearby locus in the first homologous segment in the parent is not observed in the obtained genetic data of the 60 mixed sample. In some embodiments, the method includes increasing the probability in the probability distribution of an a particular allele being present at a second locus in the mixed sample if that particular allele is present in the second homologous segment in the parent and an allele at a nearby 65 locus in the second homologous segment in the parent is observed in the obtained genetic data of the mixed sample;

12

or decreasing the probability in the probability distribution of an a particular allele being present at a second locus in the mixed sample if that particular allele is present in the second homologous segment in the parent and an allele at a nearby locus in the second homologous segment in the parent is not observed in the obtained genetic data of the mixed sample.

In some embodiments, the method includes obtaining phased genetic data for both the mother and father of the fetus. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the number of copies of the first homologous chromosome segment or portion thereof from the mother in the genome of the fetus, the number of copies of the second homologous chromosome segment or portion thereof from the mother in the genome of the fetus, the number of copies of the first homologous chromosome segment or portion thereof from the father in the genome of the fetus, the number of copies of the second homologous chromosome segment or portion thereof from the father in the genome of the fetus, and the total number of copies of the chromosome segment of interest present in the genome of the fetus. In some embodiments, the method includes calculating (such as calculating on a computer), for each of the hypotheses, a probability distribution of expected genetic data for the plurality of loci in the mixed sample from the obtained phased genetic data from the mother and father. In some embodiments, the method includes increasing the probability in the probability distribution of an a particular allele being present at a first locus in the mixed sample if that particular allele is present in the first homologous segment in the mother or father and an allele at a nearby locus in the first homologous segment in that parent is observed in the obtained genetic data of the mixed sample; or decreasing the probability in the probability distribution of an a particular allele being present at a first locus in the mixed sample if that particular allele is present in the first homologous segment in the mother or father and an allele at a nearby locus in the first homologous segment in that parent is not observed in the obtained genetic data of the mixed sample. In some embodiments, the method includes increasing the probability in the probability distribution of an a particular allele being present at a second locus in the mixed sample if that particular allele is present in the second homologous segment in the mother or father and an allele at a nearby locus in the second homologous segment in that parent is observed in the obtained genetic data of the mixed sample; or decreasing the probability in the probability distribution of an a particular allele being present at a second locus in the mixed sample if that particular allele is present in the second homologous segment in the mother or father and an allele at a nearby locus in the second homologous segment in that parent is not observed in the obtained genetic data of the mixed sample.

In some embodiments, the first locus and the locus that is nearby to the first locus co-segregate. In some embodiments, the second locus and the locus that is nearby to the second locus co-segregate. In some embodiments, no crossovers are expected to occur between the first locus and the locus that is nearby to the first locus. In some embodiments, no crossovers are expected to occur between the second locus and the locus that is nearby to the second locus. In some embodiments, the distance between the first locus and the locus that is nearby to the first locus is less than 5 mb, 1 mb, 100 kb, 10 kb, 1 kb, 0.1 kb, or 0.01 kb. In some embodiments, the distance between the second locus and the locus that is nearby to the second locus is less than 5 mb, 1 mb, 100 kb, 10 kb, 1 kb, 0.1 kb, or 0.01 kb.

## US 11,530,454 B2

1

In some embodiments, one or more crossovers occurs during the formation of a gamete that contributed a copy of the chromosome segment of interest to the fetus; and the crossover produces a chromosome segment of interest in the genome of the fetus that comprises a portion of the first 5 homologous segment and a portion of the second homologous segment from the parent. In some embodiments, the set of hypothesis comprises one or more hypotheses specifying the number of copies of the chromosome segment of interest in the genome of the fetus that comprises a portion of the 10 first homologous segment and a portion of the second homologous segment from the parent.

In some embodiments, the expected genetic data of the mixed sample comprises the expected amount of one or more of the alleles at each locus in the plurality of loci in the 15 mixed sample for each of the hypotheses.

In one aspect, the invention features a method of determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the 20 genome of an individual (such as in the genome of one or more cells, cfDNA, cfRNA, an individual suspected of having cancer, a fetus, or an embryo) using phased genetic data. In some embodiments, the method involves simultaneously or sequentially in any order (i) obtaining phased 25 genetic data for the first homologous chromosome segment comprising the identity of the allele present at that locus on the first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, (ii) obtaining phased genetic data for the 30 second homologous chromosome segment comprising the identity of the allele present at that locus on the second homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and (iii) obtaining measured genetic allelic data 3 comprising the amount of each allele at each of the loci in the set of polymorphic loci in a sample of DNA or RNA from one or more cells from the individual or in a mixed sample of cell-free DNA or RNA from two or more genetically different cells from the individual. In some embodiments, 40 the method involves calculating allele ratios for one or more loci in the set of polymorphic loci that are heterozygous in at least one cell from which the sample was derived. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles divided 45 by the total measured quantity of all the alleles for the locus. In some embodiments, the method involves determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment by comparing one or more calculated allele ratios for a locus to an expected allele 50 ratio, such as a ratio that is expected for that locus if the first and second homologous chromosome segments are present in equal proportions. In some embodiments, the expected ratio is 0.5 for biallelic loci.

In some embodiments for prenatal testing, the method 55 involves simultaneously or sequentially in any order (i) obtaining phased genetic data for the first homologous chromosome segment in the genome of a fetus (such as a fetus gestating in a pregnant mother) comprising the identity of the allele present at that locus on the first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, (ii) obtaining phased genetic data for the second homologous chromosome segment in the genome of the fetus comprising the identity of the allele present at that locus on the second 65 homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome

14

segment, and (iii) obtaining measured genetic allelic data comprising the amount of each allele at each of the loci in the set of polymorphic loci in a mixed sample of DNA or RNA from the mother of the fetus that includes fetal DNA or RNA and maternal DNA or RNA (such as a mixed sample of cell-free DNA or RNA originating from a blood sample from the mother that includes fetal cell-free DNA or RNA and maternal cell-free DNA or RNA). In some embodiments, the method involves calculating allele ratios for one or more loci in the set of polymorphic loci that are heterozygous in the fetus and/or heterozygous in the mother. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles divided by the total measured quantity of all the alleles for the locus. In some embodiments, the method involves determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment by comparing one or more calculated allele ratios for a locus to an expected allele ratio, such as a ratio that is expected for that locus if the first and second homologous chromosome segments are present in equal proportions.

In some embodiments, a calculated allele ratio is indicative of an overrepresentation of the number of copies of the first homologous chromosome segment if either (i) the allele ratio for the measured quantity of the allele present at that locus on the first homologous chromosome divided by the total measured quantity of all the alleles for the locus is greater than the expected allele ratio for that locus, or (ii) the allele ratio for the measured quantity of the allele present at that locus on the second homologous chromosome divided by the total measured quantity of all the alleles for the locus is less than the expected allele ratio for that locus. In some embodiments, a calculated allele ratio is indicative of no overrepresentation of the number of copies of the first homologous chromosome segment if either (i) the allele ratio for the measured quantity of the allele present at that locus on the first homologous chromosome divided by the total measured quantity of all the alleles for the locus is less than or equal to the expected allele ratio for that locus, or (ii) the allele ratio for the measured quantity of the allele present at that locus on the second homologous chromosome divided by the total measured quantity of all the alleles for the locus is greater than or equal to the expected allele ratio

In some embodiments, determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment. In some embodiments, predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) are estimated for each hypothesis given the degree of overrepresentation specified by that hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing the calculated allele ratios to the predicted allele ratios, and the hypothesis with the greatest likelihood is selected. In some embodiments, an expected distribution of a test statistic is calculated using the predicted allele ratios for each hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing a test statistic that is calculated using the calculated allele ratios to the expected distribution of the test statistic that is calculated using the predicted allele ratios, and the hypothesis with the greatest likelihood is selected. In some embodiments, predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are

## US 11,530,454 B2

15

heterozygous in the fetus and/or heterozygous in the mother) are estimated given the phased genetic data for the first homologous chromosome segment, the phased genetic data for the second homologous chromosome segment, and the degree of overrepresentation specified by that hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing the calculated allele ratios to the predicted allele ratios; and the hypothesis with the greatest likelihood is selected.

In some embodiments, the ratio of DNA (or RNA) from 10 one or more target cells to the total DNA (or RNA) in the sample is calculated. An exemplary ratio is the ratio of fetal DNA (or RNA) to the total DNA (or RNA) in the sample. In some embodiments, the ratio of fetal DNA to total DNA in the sample is determined by measuring the amount of an 15 allele at one or more loci in which the fetus has the allele and the mother does not have the allele. In some embodiments, the ratio of fetal DNA to total DNA in the sample is determined by measuring the difference in methylation between one or more maternal and fetal alleles. In some 20 embodiments, a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment are enumerated. In some embodiments, predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the 25 fetus and/or heterozygous in the mother) are estimated given the calculated ratio of DNA or RNA and the degree of overrepresentation specified by that hypothesis are estimated for each hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by com- 30 paring the calculated allele ratios to the predicted allele ratios, and the hypothesis with the greatest likelihood is selected. In some embodiments, an expected distribution of a test statistic calculated using the predicted allele ratios and the calculated ratio of DNA or RNA is estimated for each 3 hypothesis. In some embodiments, the likelihood that the hypothesis is correct is determined by comparing a test statistic calculated using the calculated allele ratios and the calculated ratio of DNA or RNA to the expected distribution of the test statistic calculated using the predicted allele ratios 40 and the calculated ratio of DNA or RNA, and the hypothesis with the greatest likelihood is selected.

In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome 45 segment. In some embodiments, the method includes estimating, for each hypothesis, either (i) predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) given the degree of overrepresentation 50 specified by that hypothesis or (ii) for one or more possible ratios of DNA or RNA (such as ratios of fetal DNA or RNA to the total DNA or RNA in the sample), an expected distribution of a test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA from the 55 one or more target cells (such as fetal cells) to the total DNA or RNA in the sample. In some embodiments, a data fit is calculated by comparing either (i) the calculated allele ratios to the predicted allele ratios, or (ii) a test statistic calculated using the calculated allele ratios and the possible ratio of 60 DNA or RNA to the expected distribution of the test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA. In some embodiments, one or more of the hypotheses are ranked according to the data fit, and the hypothesis that is ranked the highest is selected. In some 65 embodiments, a technique or algorithm, such as a search algorithm, is used for one or more of the following steps:

16

calculating the data fit, ranking the hypotheses, or selecting the hypothesis that is ranked the highest. In some embodiments, the data fit is a fit to a beta-binomial distribution or a fit to a binomial distribution. In some embodiments, the technique or algorithm is selected from the group consisting of maximum likelihood estimation, maximum a-posteriori estimation, Bayesian estimation, dynamic estimation (such as dynamic Bayesian estimation), and expectation-maximization estimation. In some embodiments, the method includes applying the technique or algorithm to the obtained genetic data and the expected genetic data.

In some embodiments, the method includes creating a partition of possible ratios (such as ratios of fetal DNA or RNA to the total DNA or RNA in the sample) that range from a lower limit to an upper limit for the ratio of DNA or RNA from the one or more target cells to the total DNA or RNA in the sample. In some embodiments, a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment are enumerated. In some embodiments, the method includes estimating, for each of the possible ratios of DNA or RNA in the partition and for each hypothesis, either (i) predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) given the possible ratio of DNA or RNA and the degree of overrepresentation specified by that hypothesis or (ii) an expected distribution of a test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA. In some embodiments, the method includes calculating, for each of the possible ratios of DNA or RNA in the partition and for each hypothesis, the likelihood that the hypothesis is correct by comparing either (i) the calculated allele ratios to the predicted allele ratios, or (ii) a test statistic calculated using the calculated allele ratios and the possible ratio of DNA or RNA to the expected distribution of the test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA. In some embodiments, the combined probability for each hypothesis is determined by combining the probabilities of that hypothesis for each of the possible ratios in the partition; and the hypothesis with the greatest combined probability is selected. In some embodiments, the combined probability for each hypothesis is determining by weighting the probability of a hypothesis for a particular possible ratio based on the likelihood that the possible ratio is the correct ratio.

In one aspect, the invention features a method for determining a number of copies of a chromosome or chromosome segment in the genome of one or more cells from an individual using phased or unphased genetic data. In some embodiments, the method involves obtaining genetic data at a set of polymorphic loci on the chromosome or chromosome segment in a sample by measuring the quantity of each allele at each locus. In some embodiments, the sample is a sample of DNA or RNA from one or more cells from the individual or a mixed sample of cell-free DNA from the individual that includes cell-free DNA from two or more genetically different cells. In some embodiments, allele ratios are calculated for the loci that are heterozygous in at least one cell from which the sample was derived. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles divided by the total measured quantity of all the alleles for the locus. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles (such as the allele on the first homologous chromosome segment) divided by the measured quantity of one or more other alleles (such as the allele on the second homologous

## US 11,530,454 B2

1′

chromosome segment) for the locus. In some embodiments, a set of one or more hypotheses specifying the number of copies of the chromosome or chromosome segment in the genome of one or more of the cells are enumerated. In some embodiments, the hypothesis that is most likely based on the test statistic is selected, thereby determining the number of copies of the chromosome or chromosome segment in the genome of one or more of the cells.

In one aspect, the invention features a method for determining a number of copies of a chromosome or chromosome 10 segment in the genome of a fetus (such as a fetus that is gestating in a pregnant mother) using phased or unphased genetic data. In some embodiments, the method involves obtaining genetic data at a set of polymorphic loci on the chromosome or chromosome segment in a sample by mea- 15 suring the quantity of each allele at each locus. In some embodiments, the sample is a mixed sample of DNA comprising fetal DNA or RNA and maternal DNA or RNA from the mother of the fetus (such as a mixed sample of cell-free DNA or RNA originating from a blood sample from the 20 mother that includes fetal cell-free DNA or RNA and maternal cell-free DNA or RNA). In some embodiments, allele ratios are calculated for the loci that are heterozygous in the fetus and/or heterozygous in the mother. In some embodiments, the calculated allele ratio for a particular 25 locus is the measured quantity of one of the alleles divided by the total measured quantity of all the alleles for the locus. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles (such as the allele on the first homologous chromosome 30 segment) divided by the measured quantity of one or more other alleles (such as the allele on the second homologous chromosome segment) for the locus. In some embodiments, a set of one or more hypotheses specifying the number of copies of the chromosome or chromosome segment in the 3 genome of fetus are enumerated. In some embodiments, the hypothesis that is most likely based on the test statistic is selected, thereby determining the number of copies of the chromosome or chromosome segment in the genome of the fetus.

In some embodiments, a hypotheses is selected if the probability that the test statistic belongs to a distribution of the test statistic for that hypothesis is above an upper threshold; one or more of the hypotheses is rejected if the probability that the test statistic belongs to the distribution of 45 the test statistic for that hypothesis is below an lower threshold; or a hypothesis is neither selected nor rejected if the probability that the test statistic belongs to the distribution of the test statistic for that hypothesis is between the lower threshold and the upper threshold, or if the probability 50 is not determined with sufficiently high confidence. In some embodiments, the overrepresentation of the number of copies of the first homologous chromosome segment is due to a duplication of the first homologous chromosome segment or a deletion of the second homologous chromosome seg- 55 ment. In some embodiments, the total measured quantity of all the alleles for one or more of the loci is compared to a reference amount to determine whether the overrepresentation of the number of copies of the first homologous chromosome segment is due to a duplication of the first homolo- 60 gous chromosome segment or a deletion of the second homologous chromosome segment. In some embodiments, the magnitude of the difference between the calculated allele ratio and the expected allele ratio for one or more loci is used to determine whether the overrepresentation of the number 65 of copies of the first homologous chromosome segment is due to a duplication of the first homologous chromosome

18

segment or a deletion of the second homologous chromosome segment. In some embodiments, the first and second homologous chromosome segments are determined to be present in equal proportions if there is not an overrepresentation of the number of copies of the first homologous chromosome segment, and there is not an overrepresentation of the second homologous chromosome segment (such as in the genome of the cells, cfDNA, cfRNA, individual, fetus, or embryo).

In some embodiments, the ratio of DNA from the one or more target cells to the total DNA in the sample is determined based on the total or relative amount of one or more alleles at one or more loci for which the genotype of the target cells differs from the genotype of the non-target cells and for which the target cells and non-target cells are expected to be disomic. In some embodiments, this ratio is used to determine whether the overrepresentation of the number of copies of the first homologous chromosome segment is due to a duplication of the first homologous chromosome segment or a deletion of the second homologous chromosome segment. In some embodiments, the ratio is used to determine the number of extra copies of a chromosome segment or chromosome that is duplicated. In some embodiments, the phased genetic data includes probabilistic data. In some embodiments, obtaining the phased genetic data for the first homologous chromosome segment and/or the second homologous chromosome segment in the genome of the fetus includes obtaining phased genetic data for the first homologous chromosome segment and/or the second homologous chromosome segment in the genome of one or both biological parents of the fetus, and inferring which homologous chromosome segment the fetus inherited from one or both biological parents. In some embodiments, the probability of one or more crossovers (such as 1, 2, 3, or 4 crossovers) that may have occurred during the formation of a gamete that contributed a copy of the first homologous chromosome segment or the second homologous chromosome segment to the fetus individual is used to infer which homologous chromosome segment(s) the fetus inherited from one or both biological parents. In some embodiments, phased genetic data for the mother and/or father of the fetus is obtained using a technique selected from the group consisting of digital PCR, inferring a haplotype using population based haplotype frequencies, haplotyping using a haploid cell such as a sperm or egg, haplotyping using genetic data from one or more first degree relatives, and combinations thereof. In some embodiments, the phased genetic data for the individual is obtained by phasing a portion or all of region corresponding to a deletion or duplication in a sample from the individual. In some embodiments, the phased genetic data for a fetus is obtained by phasing a portion or all of region corresponding to a deletion or duplication in a sample from the fetus or the mother of the fetus. In some embodiments, obtaining phased genetic data for the first and second homologous chromosome segments includes determining the identity of alleles present in one of the chromosome segments and determining the identity of alleles present in the other chromosome segment by inference. In some embodiments, alleles from unphased genetic data that are not present in the first homologous chromosome segment are assigned to the second homologous chromosome segment. For example, if the genotype of the individual is (AB, AB) and the phased data for the individual indicates that the first haplotype is (A,A); then, the other haplotype can be inferred to be (B,B). In some embodiments, if only one allele is measured at a locus then that allele is determined to be part of both the first and

## US 11,530,454 B2

1

second homologous chromosome segments (e.g., if the genotype is AA at a locus than both haplotypes have the A allele). In some embodiments, the phased genetic data for the individual comprises determining whether or not one or more possible chromosome crossovers occurred, such as by 5 determining the sequence of a recombination hotspot and optionally of a region flanking a recombination hotspot. In some embodiments, any of the primer libraries of the invention are used to detect a recombination event to determine what haplotype blocks are present in the genome of an 10 individual.

In some embodiments, the method includes using a joint distribution model (such as a joint distribution model that takes into account the linkage between loci), performing a linkage analysis, using a binomial distribution model, using 15 a beta-binomial distribution model, and/or using the likelihood of crossovers having occurred during the meiosis that gave rise to the gametes that formed the embryo that grew into the fetus (such as using the probability of chromosomes crossing over at different locations in a chromosome to 20 model dependence between polymorphic alleles on the chromosome or chromosome segment of interest).

In some embodiments, one or more of the calculated allele ratios for the cfDNA or cfRNA are indicative of the corresponding allele ratios for DNA or RNA in the cells from 25 which the cfDNA or cfRNA was derived. In some embodiments, one or more of the calculated allele ratios for the cfDNA or cfRNA are indicative of the corresponding allele ratios in the genome of the individual. In some embodiments, an allele ratio is only calculated or is only compared 30 to an expected allele ratio if the measured genetic data indicate that more than one different allele is present for that locus in the sample (such as in a cfDNA or cfRNA sample). In some embodiments, an allele ratio is only calculated or is only compared to an expected allele ratio if the locus is 3 heterozygous in at least one of the cells from which the sample was derived (such as a locus that is heterozygous in the fetus and/or heterozygous in the mother). In some embodiments, an allele ratio is only calculated or is only compared to an expected allele ratio if the locus is heterozy- 40 gous in the fetus. In some embodiments, an allele ratio is calculated and compared to an expected allele ratio for a homozygous locus. For example, allele ratios for loci that are predicted to be homozygous for a particular individual being tested (or for both a fetus and pregnant mother) may 45 be analyzed to determine the level of noise or error in the system.

In some embodiments, at least 10; 50; 100; 200; 300; 500; 750; 1,000; 2,000; 3,000; 4,000, or more loci (such as SNPs) are analyzed for a chromosome or chromosome segment of 50 interest. In some embodiments, the average number of loci (such as SNPs) per mb in a chromosome or chromosome segment of interest is at least 1; 10; 25; 50; 100; 150; 200; 300; 500; 750; 1,000; or more loci per mb. In some embodiments, the average number of loci (such as SNPs) per mb in 55 a chromosome or chromosome segment of interest is between 1 and 500 loci per mb, such as between 1 and 50, 50 and 100, 100 and 200, 200 and 400, 200 and 300, or 300 and 400 loci per mb, inclusive. In some embodiments, loci in multiple portions of a potential deletion or duplication are 60 analyzed to increase the sensitivity and/or specificity of the CNV determination compared to only analyzing 1 loci or only analyzing a few loci that are near each other. In some embodiments, only the two most common alleles at each locus are measured or are used to determine the calculated 65 allele ratio. In some embodiments, the amplification of loci is performed using a polymerase (e.g., a DNA polymerase,

20

RNA polymerase, or reverse transcriptase) with low 5'→3' exonuclease and/or low strand displacement activity. In some embodiments, the measured genetic allelic data is obtained by (i) sequencing the DNA or RNA in the sample, (ii) amplifying DNA or RNA in the sample and then sequencing the amplified DNA, or (ii) amplifying the DNA or RNA in the sample, ligating PCR products, and then sequencing the ligated products. In some embodiments, measured genetic allelic data is obtained by dividing the DNA or RNA from the sample into a plurality of fractions, adding a different barcode to the DNA or RNA in each fraction (e.g., such that all the DNA or RNA in a particular fraction has the same barcode), optionally amplifying the barcoded DNA or RNA, combining the fractions, and then sequencing the barcoded DNA or RNA in the combined fractions. In some embodiments, alleles of the polymorphic loci (such as SNPs) are identified using one or more of the following methods: sequencing (such as nanopore sequencing or Halcyon Molecular sequencing), SNP array, real time PCR, TaqMan, Nanostring nCounter® Analysis System, Illumina GoldenGate Genotyping Assay that uses a discriminatory DNA polymerase and ligase, ligation-mediated PCR, or Linked Inverted Probes (LIPs; which can also be called pre-circularized probes, pre-circularizing probes, circularizing probes, Padlock Probes, or Molecular Inversion Probes (MIPs)). In some embodiments, two or more (such as 3 or 4) target amplicons are ligated together and then the ligated products are sequenced. In some embodiments, measurements for different alleles for the same locus are adjusted for differences in metabolism, apoptosis, histones, inactivation, and/or amplification between the alleles (such as differences in amplification efficiency between different alleles of the same locus). In some embodiments, this adjustment is performed prior to calculating allele ratios for the obtained genetic data or prior to comparing the measured genetic data to the expected genetic data.

In some embodiments, the method also includes determining the presence or absence of one or more risk factors for a disease or disorder. In some embodiments, the method also includes determining the presence or absence of one or more polymorphisms or mutations associated with the disease or disorder or an increased risk for a disease or disorder. In some embodiments, the method also includes determining the total level of cfDNA cf mDNA, cf nDNA, cfRNA, miRNA, or any combination thereof. In some embodiments, the method includes determining the level of one or more cfDNA cf mDNA, cf nDNA, cfRNA, and/or miRNA molecules of interest, such as molecules with a polymorphism or mutation associated with a disease or disorder or an increased risk for a disease or disorder. In some embodiments, the fraction of tumor DNA out of total DNA (such as the fraction of tumor cfDNA out of total cfDNA or the fraction of tumor cfDNA with a particular mutation out of total cfDNA) is determined. In some embodiments, this tumor fraction is used to determine the stage of a cancer (since higher tumor fractions can be associated with more advanced stages of cancer). In some embodiments, the method also includes determining the total level of DNA or RNA level. In some embodiments, the method includes determining the methylation level of one or more DNA or RNA molecules of interest, such as molecules with a polymorphism or mutation associated with a disease or disorder or an increased risk for a disease or disorder. In some embodiments, the method includes determining the presence or absence of a change in DNA integrity. In some embodiments, the method also includes determining the total level of mRNA splicing. In some embodiments, the method

## US 11,530,454 B2

2

includes determining the level of mRNA splicing or detecting alternative mRNA splicing for one or RNA molecules of interest, such as molecules with a polymorphism or mutation associated with a disease or disorder or an increased risk for a disease or disorder.

In some embodiments, the invention features a method for detecting a cancer phenotype in an individual, wherein the cancer phenotype is defined by the presence of at least one of a set of mutations. In some embodiments, the method includes obtaining DNA or RNA measurements for a sample 10 of DNA or RNA from one or more cells from the individual, wherein one or more of the cells is suspected of having the cancer phenotype; and analyzing the DNA or RNA measurements to determine, for each of the mutations in the set of mutations, the likelihood that at least one of the cells has 15 that mutation. In some embodiments, the method includes determining that the individual has the cancer phenotype if either (i) for at least one of the mutations, the likelihood that at least one of the cells contains that mutations is greater than a threshold, or (ii) for at least one of the mutations, the 20 likelihood that at least one of the cells has that mutations is less than the threshold, and for a plurality of the mutations, the combined likelihood that at least one of the cells has at least one of the mutations is greater than the threshold. In some embodiments, one or more cells have a subset or all of 25 the mutations in the set of mutations. In some embodiments, the subset of mutations is associated with cancer or an increased risk for cancer. In some embodiments, the sample includes cell-free DNA or RNA. In some embodiments, the DNA or RNA measurements include measurements (such as 30 the quantity of each allele at each locus) at a set of polymorphic loci on one or more chromosomes or chromosome segments of interest.

In one aspect, the invention features methods for selecting a therapy for the treatment, stabilization, or prevention of a 35 disease or disorder in a mammal. In some embodiments, the method includes determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment using any of the methods described herein. In some embodiments, a therapy is selected for the mammal (such as a therapy for a disease or disorder associated with the overrepresentation of the first homologous chromosome segment)

In one aspect, the invention features methods for preventing, delaying, stabilizing, or treating a disease or disorder in a mammal. In some embodiments, the method includes determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment susing any of the methods described herein. In some embodiments, a therapy is selected for the mammal (such as a therapy for a disease or disorder associated with the overrepresentation of the first homologous chromosome segment) and then the therapy is administered to the mammal.

In some embodiments, treating, stabilizing, or preventing a disease or disorder includes preventing or delaying an initial or subsequent occurrence of a disease or disorder, increasing the disease-free survival time between the disappearance of a condition and its reoccurrence, stabilizing or reducing an adverse symptom associated with a condition, or inhibiting or stabilizing the progression of a condition. In some embodiments, at least 20, 40, 60, 80, 90, or 95% of the treated subjects have a complete remission in which all evidence of the condition disappears. In some embodiments, 65 the length of time a subject survives after being diagnosed with a condition and treated is at least 20, 40, 60, 80, 100,

22

200, or even 500% greater than (i) the average amount of time an untreated subject survives or (ii) the average amount of time a subject treated with another therapy survives.

In some embodiments, treating, stabilizing, or preventing cancer includes reducing or stabilizing the size of a tumor (e.g., a benign or malignant tumor), slowing or preventing an increase in the size of a tumor, reducing or stabilizing the number of tumor cells, increasing the disease-free survival time between the disappearance of a tumor and its reappearance, preventing an initial or subsequent occurrence of a tumor, or reducing or stabilizing an adverse symptom associated with a tumor. In one embodiment, the number of cancerous cells surviving the treatment is at least 10, 20, 40, 60, 80, or 100% lower than the initial number of cancerous cells, as measured using any standard assay. In some embodiments, the decrease in the number of cancerous cells induced by administration of a therapy of the invention is at least 2, 5, 10, 20, or 50-fold greater than the decrease in the number of non-cancerous cells. In some embodiments, the number of cancerous cells present after administration of a therapy is at least 2, 5, 10, 20, or 50-fold lower than the number of cancerous cells present after administration of a control (such as administration of saline or a buffer). In some embodiments, the methods of the present invention result in a decrease of 10, 20, 40, 60, 80, or 100% in the size of a tumor as determined using standard methods. In some embodiments, at least 10, 20, 40, 60, 80, 90, or 95% of the treated subjects have a complete remission in which there are no detectable cancerous cells. In some embodiments, the cancer does not reappear, or reappears after at least 2, 5, 10, 15, or 20 years. In some embodiments, the length of time a subject survives after being diagnosed with cancer and treated with a therapy of the invention is at least 10, 20, 40, 60, 80, 100, 200, or even 500% greater than (i) the average amount of time an untreated subject survives or (ii) the average amount of time a subject treated with another therapy survives.

In one aspect, the invention features methods for stratification of subjects involved in a clinical trial for the treatment, stabilization, or prevention of a disease or disorder in a mammal. In some embodiments, the method includes determining if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment using any of the methods described herein before, during, or after the clinical trial. In some embodiments, the presence or absence of the overrepresentation of the first homologous chromosome segment in the genome of the subject places the subject into a subgroup for the clinical trial.

In some embodiments, the disease or disorder is selected from the group consisting of cancer, mental handicap, learning disability (e.g., idiopathic learning disability), mental retardation, developmental delay, autism, neurodegenerative disease or disorder, schizophrenia, physical handicap, autoimmune disease or disorder, systemic lupus erythematosus, psoriasis, Crohn's disease, glomerulonephritis, HIV infection, AIDS, and combinations thereof. In some embodiments, the disease or disorder is selected from the group consisting of DiGeorge syndrome, DiGeorge 2 syndrome, DiGeorge/VCFS syndrome, Prader-Willi syndrome, Angelman syndrome, Beckwith-Wiedemann syndrome, 1p36 deletion syndrome, 2q37 deletion syndrome, 3q29 deletion syndrome, 9q34 deletion syndrome, 17q21.31 deletion syndrome, Cri-du-chat syndrome, Jacobsen syndrome, Miller Dieker syndrome, Phelan-McDermid syndrome, Smith-Magenis syndrome, WAGR syndrome, Wolf-Hirschhorn syn-

## US 11,530,454 B2

23

drome, Williams syndrome, Williams-Beuren syndrome, Miller-Dieker syndrome, Phelan-McDermid syndrome, Smith-Magenis syndrome, Down syndrome, Edward syndrome, Patau syndrome, Klinefelter syndrome, Turner syndrome, 47,XXX syndrome, 47,XYY syndrome, Sotos syndrome, and combinations thereof. In some embodiments, the method determines the presence or absence of one or more of the following chromosomal abnormalities: null somy, monosomy, uniparental di somy, trisomy, matched trisomy, unmatched trisomy, maternal trisomy, paternal trisomy, trip- 10 loidy, mosaicism tetrasomy, matched tetrasomy, unmatched tetrasomy, other aneuploidies, unbalanced translocations, balanced translocations, insertions, deletions, recombinations, and combinations thereof. In some embodiments, the chromosomal abnormality is any deviation in the copy 15 number of a specific chromosome or chromosome segment from the most common number of copies of that segment or chromosome, for example in a human somatic cell, any deviation from 2 copies can be regarded as a chromosomal abnormality. In some embodiments, the method determines 20 the presence or absence of a euploidy. In some embodiments, the copy number hypotheses include one or more copy number hypotheses for a singleton pregnancy. In some embodiments, the copy number hypotheses include one or more copy number hypotheses for a multiple pregnancy, 25 such as a twin pregnancy (e.g., identical or fraternal twins or a vanishing twin). In some embodiments, the copy number hypotheses include all fetuses in a multiple pregnancy being euploid, all fetuses in a multiple pregnancy being aneuploid (such as any of the aneuploidies disclosed herein), and/or 30 one or more fetuses in a multiple pregnancy being euploid and one or more fetuses in a multiple pregnancy being aneuploidy. In some embodiments, the copy number hypotheses include identical twins (also referred to as monozygotic twins) or fraternal twins (also referred to as dizygotic twins). 3: In some embodiments, the copy number hypotheses include a molar pregnancy, such as a complete or partial molar pregnancy. In some embodiments, the chromosome segment of interest is an entire chromosome. In some embodiments. the chromosome or chromosome segment is selected from 40 the group consisting of chromosome 13, chromosome 18, chromosome 21, the X chromosome, the Y chromosome, segments thereof, and combinations thereof. In some embodiments, the first homologous chromosome segment and second homologous chromosome segment are a pair of 45 homologous chromosome segments that comprises the chromosome segment of interest. In some embodiments, the first homologous chromosome segment and second homologous chromosome segment are a pair of homologous chromosomes of interest. In some embodiments, a confidence is 50 computed for the CNV determination or the diagnosis of the disease or disorder.

In some embodiments, the deletion is a deletion of at least 0.01 kb, 0.1 kb, 1 kb, 10 kb, 100 kb, 1 mb, 2 mb, 3 mb, 5 mb, 10 mb, 15 mb, 20 mb, 30 mb, or 40 mb. In some 55 embodiments, the deletion is a deletion of between 1 kb to 40 mb, such as between 1 kb to 100 kb, 100 kb to 1 mb, 1 to 5 mb, 5 to 10 mb, 10 to 15 mb, 15 to 20 mb, 20 to 25 mb, 25 to 30 mb, or 30 to 40 mb, inclusive. In some embodiments, one copy of the chromosome segment is deleted and one copy is present. In some embodiments, two copies of the chromosome segment are deleted. In some embodiments, an entire chromosome is deleted.

In some embodiments, the duplication is a duplication of at least 0.01 kb, 0.1 kb, 1 kb, 10 kb, 100 kb, 1 mb, 2 mb, 3 65 mb, 5 mb, 10 mb, 15 mb, 20 mb, 30 mb, or 40 mb. In some embodiments, the duplication is a duplication of between 1

24

kb to 40 mb, such as between 1 kb to 100 kb, 100 kb to 1 mb, 1 to 5 mb, 5 to 10 mb, 10 to 15 mb, 15 to 20 mb, 20 to 25 mb, 25 to 30 mb, or 30 to 40 mb, inclusive. In some embodiments, the chromosome segment is duplicated one time. In some embodiments, the chromosome segment is duplicated more than one time, such as 2, 3, 4, or 5 times. In some embodiments, an entire chromosome is duplicated. In some embodiments, a region in a first homologous segment is deleted, and the same region or another region in the second homologous segment is duplicated. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 98, 99, or 100% of the SNVs tested for are transversion mutations rather than transition mutations.

In some embodiments, the sample comprises DNA and/or RNA from (i) one or more target cells or (ii) one or more non-target cells. In some embodiments, the sample is a mixed sample with DNA and/or RNA from one or more target cells and one or more non-target cells. In some embodiments, the target cells are cells that have a CNV, such as a deletion or duplication of interest, and the non-target cells are cells that do not have the copy number variation of interest. In some embodiments in which the one or more target cells are cancer cell(s) and the one or more non-target cells are non-cancerous cell(s), the method includes determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more of the cancer cells. In some embodiments in which the one or more target cells are genetically identical cancer cell(s) and the one or more non-target cells are non-cancerous cell(s), the method includes determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment in the genome of the cancer cell(s). In some embodiments in which the one or more target cells are genetically non-identical cancer cell(s) and the one or more non-target cells are non-cancerous cell(s), the method includes determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more of the genetically non-identical cancer cells. In some embodiments in which the sample comprises cell-free DNA from a mixture of one or more cancer cells and one or more non-cancerous cells, the method includes determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more of the cancer cells. In some embodiments in which the one or more target cells are genetically identical fetal cell(s) and the one or more non-target cells are maternal cell(s), the method includes determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment in the genome of the fetal cell(s). In some embodiments in which the one or more target cells are genetically non-identical fetal cell(s) and the one or more non-target cells are maternal cell(s), the method includes determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more of the genetically non-identical fetal cells. As the cells of most individuals contain a nearly identical set of nuclear DNA, the term "target cell" may be used interchangeably with the term "individual" in some embodiments. Cancerous cells have genotypes that are distinct from the host individual. In this case, the cancer itself may be considered an individual. Moreover, many cancers are heterogeneous meaning that different cells in a tumor are genetically distinct from other cells in the same tumor. In this case, the different genetically identical regions can be considered different individuals. Alternately, the cancer may

## US 11,530,454 B2

2

be considered a single individual with a mixture of cells with distinct genomes. Typically, non-target cells are euploid, though this is not necessarily the case.

In some embodiments, the sample is obtained from a maternal whole blood sample or fraction thereof, cells isolated from a maternal blood sample, an amniocentesis sample, a products of conception sample, a placental tissue sample, a chorionic villus sample, a placental membrane sample, a cervical mucus sample, or a sample from a fetus. In some embodiments, the sample comprises cell-free DNA 10 obtained from a blood sample or fraction thereof from the mother. In some embodiments, the sample comprises nuclear DNA obtained from a mixture of fetal cells and maternal cells. In some embodiments, the sample is obtained from a fraction of maternal blood containing nucleated cells 15 that has been enriched for fetal cells. In some embodiments, a sample is divided into multiple fractions (such as 2, 3, 45, or more fractions) that are each analyzed using a method of the invention. If each fraction produces the same results (such as the presence or absence of one or more CNVs of 20 interest), the confidence in the results increases. In different fractions produce different results, the sample could be re-analyzed or another sample could be collected from the same subject and analyzed.

Exemplary subjects include mammals, such as humans 25 and mammals of veterinary interest. In some embodiments, the mammal is a primate (e.g., a human, a monkey, a gorilla, an ape, a lemur, etc.), a bovine, an equine, a porcine, a canine, or a feline.

In some embodiments, any of the methods include generating a report (such as a written or electronic report) disclosing a result of the method of the invention (such as the presence or absence of a deletion or duplication).

In some embodiments, any of the methods include taking a clinical action based on a result of a method of the 35 invention (such as the presence or absence of a deletion or duplication). In some embodiments in which an embryo or fetus has one or more polymorphisms or mutations of interest (such as a CNV) based on a result of a method of the invention, the clinical action includes performing additional 40 testing (such as testing to confirm the presence of the polymorphism or mutation), not implanting the embryo for IVF, implanting a different embryo for IVF, terminating a pregnancy, preparing for a special needs child, or undergoing an intervention designed to decrease the severity of the 45 phenotypic presentation of a genetic disorder. In some embodiments, the clinical action is selected from the group consisting of performing an ultrasound, amniocentesis on the fetus, amniocentesis on a subsequent fetus that inherits genetic material from the mother and/or father, chorion 50 villus biopsy on the fetus, chorion villus biopsy on a subsequent fetus that inherits genetic material from the mother and/or father, in vitro fertilization, preimplantation genetic diagnosis on one or more embryos that inherited genetic material from the mother and/or father, karyotyping 55 on the mother, karyotyping on the father, fetal echocardiogram (such as an echocardiogram of a fetus with trisomy 21, 18, or 13, monosomy X, or a microdeletion) and combinations thereof. In some embodiments, the clinical action is selected from the group consisting of administering growth 60 hormone to a born child with monosomy X (such as administration starting at ~9 months), administering calcium to a born child with a 22q deletion (such as DiGeorge syndrome), administering an androgen such as testosterone to a born child with 47,XXY (such as one injection per month for 65 3 months of 25 mg testosterone enanthate to an infant or toddler), performing a test for cancer on a woman with a

26

complete or partial molar pregnancy (such as a triploid fetus), administering a therapy for cancer such as a chemotherapeutic agent to a woman with a complete or partial molar pregnancy (such as a triploid fetus), screening a fetus determined to be male (such as a fetus determined to be male using a method of the invention) for one or more X-linked genetic disorders such as Duchenne muscular dystrophy (DMD), adrenoleukodystrophy, or hemophilia, performing amniocentesis on a male fetus at risk for an X-linked disorder, administering dexamethasone to a women with a female fetus (such as a fetus determined to be female using a method of the invention) at risk for congenital adrenal hyperplasia, performing amniocentesis on a female fetus at risk for congenital adrenal hyperplasia, administering killed vaccines (instead of live vaccines) or not administering certain vaccines to a born child that is (or is suspected of being) immune deficient from a 22q11.2 deletion, performing occupational and/or physical therapy, performing early intervention in education, delivering the baby at a tertiary care center with a NICU and/or having pediatric specialists available at delivery, behavioral intervention for born child (such as a child with XXX, XXY, or XYY), and combina-

In some embodiments, ultrasound or another screening test is performed on a women determined to have multiple pregnancies (such as twins) to determine whether or not two or more of the fetus are monochorionic. Monozygotic twins result from ovulation and fertilization of a single oocyte, with subsequent division of the zygote; placentation may be dichorionic or monochorionic. Dizygotic twins occur from ovulation and fertilization of two oocytes, which usually results in dichorionic placentation. Monochorionic twins have a risk of twin-to-twin transfusion syndrome, which may cause unequal distribution of blood between fetuses that results in differences in their growth and development, sometimes resulting in stillbirth. Thus, twins determined to be monozygotic twins using a method of the invention are desirably tested (such as by ultrasound) to determine if they are monochorionic twins, and if so, these twins can be monitored (such as bi-weekly ultrasounds from 16 weeks) for signs of win-to-twin transfusion syndrome.

In some embodiments in which an embryo or fetus does not have one or more one or more polymorphisms or mutations of interest (such as a CNV) based on a result of a method of the invention, the clinical action includes implanting the embryo for IVF or continuing a pregnancy. In some embodiments, the clinical action is additional testing to confirm the absence of the polymorphism or mutation selected from the group consisting of performing an ultrasound, amniocentesis, chorion villus biopsy, and combinations thereof.

In some embodiments in which an individual has one or more one or more polymorphisms or mutations (such as a polymorphism or mutation associated with a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer) based on a result of a method of the invention, the clinical action includes performing additional testing or administering one or more therapies for a disease or disorder (such as a therapy for cancer, a therapy for the specific type of cancer or type of mutation the individual is diagnosed with, or any of the therapies disclosed herein). In some embodiments, the clinical action is additional testing to confirm the presence or absence of a polymorphism or mutation selected from the group consisting of biopsy, surgery, medical imaging (such as a mammogram or an ultrasound), and combinations thereof.

## US 11,530,454 B2

2

In some embodiments, the additional testing includes performing the same or a different method (such as any of the methods described herein) to confirm the presence or absence of the polymorphism or mutation (such as a CNV), such as testing either a second fraction of the same sample that was tested or a different sample from the same individual (such as the same pregnant mother, fetus, embryo, or individual at increased risk for cancer). In some embodiments, the additional testing is performed for an individual for whom the probability of a polymorphism or mutation 10 (such as a CNV) is above a threshold value (such as additional testing to confirm the presence of a likely polymorphism or mutation). In some embodiments, the additional testing is performed for an individual for whom the confidence or z-score for the determination of a polymor- 15 phism or mutation (such as a CNV) is above a threshold value (such as additional testing to confirm the presence of a likely polymorphism or mutation). In some embodiments, the additional testing is performed for an individual for whom the confidence or z-score for the determination of a 20 polymorphism or mutation (such as a CNV) is between minimum and maximum threshold values (such as additional testing to increase the confidence that the initial result is correct). In some embodiments, the additional testing is performed for an individual for whom the confidence for the 25 determination of the presence or absence of a polymorphism or mutation (such as a CNV) is below a threshold value (such as a "no call" result due to not being able to determine the presence or absence of the CNV with sufficient confidence). An exemplary Z core is calculated in Chiu et al. BMJ 30 2011; 342:c7401 (which is hereby incorporated by reference in its entirety) in which chromosome 21 is used as an example and can be replaced with any other chromosome or chromosome segment in the test sample.

Z score for percentage chromosome 21 in test case— ((percentage chromosome 21 in test case)— (mean percentage chromosome 21 in reference controls))/(standard deviation of percentage chromosome 21 in reference controls).

In some embodiments, the additional testing is performed 40 for an individual for whom the initial sample did not meet quality control guidelines or had a fetal fraction or a tumor fraction below a threshold value. In some embodiments, the method includes selecting an individual for additional testing based on the result of a method of the invention, the 45 probability of the result, the confidence of the result, or the z-score; and performing the additional testing on the individual (such as on the same or a different sample). In some embodiments, a subject diagnosed with a disease or disorder (such as cancer) undergoes repeat testing using a method of 50 the invention or known testing for the disease or disorder at multiple time points to monitor the progression of the disease or disorder or the remission or reoccurrence of the disease or disorder.

In one aspect, the invention features a report (such as a 55 written or electronic report) with a result from a method of the invention (such as the presence or absence of a deletion or duplication).

In various embodiments, the primer extension reaction or the polymerase chain reaction includes the addition of one or 60 more nucleotides by a polymerase. In some embodiments, the primers are in solution. In some embodiments, the primers are in solution and are not immobilized on a solid support. In some embodiments, the primers are not part of a microarray. In various embodiments, the primer extension 65 reaction or the polymerase chain reaction does not include ligation-mediated PCR. In various embodiments, the primer

28

extension reaction or the polymerase chain reaction does not include the joining of two primers by a ligase. In various embodiments, the primers do not include Linked Inverted Probes (LIPs), which can also be called pre-circularized probes, pre-circularizing probes, circularizing probes, Padlock Probes, or Molecular Inversion Probes (MIPs).

It is understood that aspects and embodiments of the invention described herein include combinations of any two or more of the aspects or embodiments of the invention.

#### Definitions

Single Nucleotide Polymorphism (SNP) refers to a single nucleotide that may differ between the genomes of two members of the same species. The usage of the term should not imply any limit on the frequency with which each variant occurs.

Sequence refers to a DNA sequence or a genetic sequence. It may refer to the primary, physical structure of the DNA molecule or strand in an individual. It may refer to the sequence of nucleotides found in that DNA molecule, or the complementary strand to the DNA molecule. It may refer to the information contained in the DNA molecule as its representation in silico.

Locus refers to a particular region of interest on the DNA of an individual, which may refer to a SNP, the site of a possible insertion or deletion, or the site of some other relevant genetic variation. Disease-linked SNPs may also refer to disease-linked loci.

Polymorphic Allele, also "Polymorphic Locus," refers to an allele or locus where the genotype varies between individuals within a given species. Some examples of polymorphic alleles include single nucleotide polymorphisms, short 35 tandem repeats, deletions, duplications, and inversions.

Polymorphic Site refers to the specific nucleotides found in a polymorphic region that vary between individuals.

controls))/(standard deviation of percentage chromosome 21 in reference controls).

Mutation refers to an alteration in a naturally-occurring or reference nucleic acid sequence, such as an insertion, deletion, duplication, translocation, substitution, frameshift mutation point mutation, ronsense mutation, missense mutation, point mutation, transition mutation, transversion mutation, reverse mutation, or microsatellite alteration. In some embodiments, the amino acid sequence encoded by the nucleic acid sequence has at least one amino acid alteration from a naturally-occurring or reference nucleic acid sequence, such as an insertion, deletion, duplication, translocation, substitution, frameshift mutation, point mutation, transition mutation, reverse mutation, or microsatellite alteration. In some embodiments, the amino acid sequence encoded by the nucleic acid sequence has at least one amino acid alteration from a naturally-occurring or reference nucleic acid sequence, such as an insertion, deletion, duplication, translocation, substitution, frameshift mutation, point mutation, reverse mutation, or microsatellite alteration. In some embodiments, the amino acid sequence encoded by the nucleic acid sequence, such as an insertion, deletion, duplication, translocation, substitution, frameshift mutation, point mutation, reverse mutation, or microsatellite alteration. In some embodiments, the amino acid sequence encoded by the nucleic acid sequence and alteration in a naturally-occurring or reference nucleic acid sequence, such as an insertion, deletion, duplication, translocation, substitution, frameshift mutation, point mutation, reverse mutation, or microsatellite alteration. In some embodiments, the amino acid sequence encoded by the nucleic acid sequence has at least one amino acid alteration from a naturally-occurring sequence.

Allele refers to the genes that occupy a particular locus. Genetic Data also "Genotypic Data" refers to the data describing aspects of the genome of one or more individuals. It may refer to one or a set of loci, partial or entire sequences, partial or entire chromosomes, or the entire genome. It may refer to the identity of one or a plurality of nucleotides; it may refer to a set of sequential nucleotides, or nucleotides from different locations in the genome, or a combination thereof. Genotypic data is typically in silico, however, it is also possible to consider physical nucleotides in a sequence as chemically encoded genetic data. Genotypic Data may be said to be "on," "of," "at," "from" or "on" the individual(s). Genotypic Data may refer to output measurements from a genotyping platform where those measurements are made on genetic material.

Genetic Material also "Genetic Sample" refers to physical matter, such as tissue or blood, from one or more individuals comprising DNA or RNA.

Confidence refers to the statistical likelihood that the called SNP, allele, set of alleles, determined number of copies of a chromosome or chromosome segment, or diag-

nosis of the presence or absence of a disease correctly represents the real genetic state of the individual.

Ploidy Calling, also "Chromosome Copy Number Calling," or "Copy Number Calling" (CNC), may refer to the act of determining the quantity and/or chromosomal identity of one or more chromosomes or chromosome segments present in a cell.

Aneuploidy refers to the state where the wrong number of chromosomes (e.g., the wrong number of full chromosomes or the wrong number of chromosome segments, such as the presence of deletions or duplications of a chromosome segment) is present in a cell. In the case of a somatic human cell it may refer to the case where a cell does not contain 22 pairs of autosomal chromosomes and one pair of sex chromosomes. In the case of a human gamete, it may refer to the case where a cell does not contain one of each of the 23 chromosomes. In the case of a single chromosome type, it may refer to the case where more or less than two homologous but non-identical chromosome copies are present, or where there are two chromosome copies present that originate from the same parent. In some embodiments, the deletion of a chromosome segment is a microdeletion.

Ploidy State refers to the quantity and/or chromosomal identity of one or more chromosomes or chromosome segments in a cell.

Chromosome may refer to a single chromosome copy, meaning a single molecule of DNA of which there are 46 in a normal somatic cell; an example is 'the maternally derived chromosome 18'. Chromosome may also refer to a chromosome type, of which there are 23 in a normal human somatic 30 cell; an example is 'chromosome 18'.

Chromosomal Identity may refer to the referent chromosome number, i.e. the chromosome type. Normal humans have 22 types of numbered autosomal chromosome types, and two types of sex chromosomes. It may also refer to the 3 parental origin of the chromosome. It may also refer to a specific chromosome inherited from the parent. It may also refer to other identifying features of a chromosome.

Allelic Data refers to a set of genotypic data concerning a set of one or more alleles. It may refer to the phased, 40 haplotypic data. It may refer to SNP identities, and it may refer to the sequence data of the DNA, including insertions, deletions, repeats and mutations. It may include the parental origin of each allele.

Allelic State refers to the actual state of the genes in a set 45 of one or more alleles. It may refer to the actual state of the genes described by the allelic data.

Allele Count refers to the number of sequences that map to a particular locus, and if that locus is polymorphic, it refers to the number of sequences that map to each of the 50 alleles. If each allele is counted in a binary fashion, then the allele count will be whole number. If the alleles are counted probabilistically, then the allele count can be a fractional number.

Allele Count Probability refers to the number of 55 sequences that are likely to map to a particular locus or a set of alleles at a polymorphic locus, combined with the probability of the mapping. Note that allele counts are equivalent to allele count probabilities where the probability of the mapping for each counted sequence is binary (zero or one). In some embodiments, the allele count probabilities may be binary. In some embodiments, the allele count probabilities may be set to be equal to the DNA measurements.

Allelic Distribution, or "allele count distribution" refers to the relative amount of each allele that is present for each 65 locus in a set of loci. An allelic distribution can refer to an individual, to a sample, or to a set of measurements made on 30

a sample. In the context of digital allele measurements such as sequencing, the allelic distribution refers to the number or probable number of reads that map to a particular allele for each allele in a set of polymorphic loci. In the context of analog allele measurements such as SNP arrays, the allelic distribution refers to allele intensities and/or allele ratios. The allele measurements may be treated probabilistically, that is, the likelihood that a given allele is present for a give sequence read is a fraction between 0 and 1, or they may be treated in a binary fashion, that is, any given read is considered to be exactly zero or one copies of a particular allele.

Allelic Distribution Pattern refers to a set of different allele distributions for different contexts, such as different parental contexts. Certain allelic distribution patterns may be indicative of certain ploidy states.

Allelic Bias refers to the degree to which the measured ratio of alleles at a heterozygous locus is different to the ratio that was present in the original sample of DNA or RNA. The degree of allelic bias at a particular locus is equal to the observed allelic ratio at that locus, as measured, divided by the ratio of alleles in the original DNA or RNA sample at that locus. Allelic bias maybe due to amplification bias, purification bias, or some other phenomenon that affects different alleles differently.

Allelic imbalance refers for SNVs, to the proportion of abnormal DNA is typically measured using mutant allele frequency (number of mutant alleles at a locus/total number of alleles at that locus). Since the difference between the amounts of two homologs in tumours is analogous, we measure the proportion of abnormal DNA for a CNV by the average allelic imbalance (AAI), defined as |(H1-H2)|/(H1+H2), where Hi is the average number of copies of homolog i in the sample and Hi/(H1+H2) is the fractional abundance, or homolog ratio, of homolog i. The maximum homolog ratio is the homolog ratio of the more abundant homolog.

Assay drop-out rate is the percentage of SNPs with no reads, estimated using all SNPs.

Single allele drop-out (ADO) rate is the percentage of SNPs with only one allele present, estimated using only heterozygous SNPs.

Primer, also "PCR probe" refers to a single nucleic acid molecule (such as a DNA molecule or a DNA oligomer) or a collection of nucleic acid molecules (such as DNA molecules or DNA oligomers) where the molecules are identical, or nearly so, and wherein the primer contains a region that is designed to hybridize to a targeted locus (e.g., a targeted polymorphic locus or a non-polymorphic locus) or to a universal priming sequence, and may contain a priming sequence designed to allow PCR amplification. A primer may also contain a molecular barcode. A primer may contain a random region that differs for each individual molecule.

Library of primers refers to a population of two or more primers. In various embodiments, the library includes at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primers. In various embodiments, the library includes at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different primer pairs, wherein each pair of primers includes a forward test primer and a reverse test primer where each pair of test primers hybridize to a target locus. In some embodiments, the library of primers includes at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different individual primers that each hybridize to a different target locus, wherein the individual primers are not

part of primer pairs. In some embodiments, the library has both (i) primer pairs and (ii) individual primers (such as universal primers) that are not part of primer pairs.

Different primers refers to non-identical primers.

Different pools refers to non-identical pools.

Different target loci refers to non-identical target loci. Different amplicons refers to non-identical amplicons.

Hybrid Capture Probe refers to any nucleic acid sequence, possibly modified, that is generated by various methods such as PCR or direct synthesis and intended to be complementary to one strand of a specific target DNA sequence in a sample. The exogenous hybrid capture probes may be added to a prepared sample and hybridized through a denature-reannealing process to form duplexes of exogenous-endogenous fragments. These duplexes may then be physically 15 separated from the sample by various means.

Sequence Read refers to data representing a sequence of nucleotide bases that were measured, e.g., using a clonal sequencing method. Clonal sequencing may produce sequence data representing single, or clones, or clusters of 20 one original DNA molecule. A sequence read may also have associated quality score at each base position of the sequence indicating the probability that nucleotide has been called correctly.

Mapping a sequence read is the process of determining a 25 sequence read's location of origin in the genome sequence of a particular organism. The location of origin of sequence reads is based on similarity of nucleotide sequence of the read and the genome sequence.

Matched Copy Error, also "Matching Chromosome Aneuploidy" (MCA), refers to a state of aneuploidy where one cell contains two identical or nearly identical chromosomes. This type of aneuploidy may arise during the formation of the gametes in meiosis, and may be referred to as a meiotic non-disjunction error. This type of error may arise in mitosis. 35 Matching trisomy may refer to the case where three copies of a given chromosome are present in an individual and two of the copies are identical.

Unmatched Copy Error, also "Unique Chromosome Aneuploidy" (UCA), refers to a state of aneuploidy where 40 one cell contains two chromosomes that are from the same parent, and that may be homologous but not identical. This type of aneuploidy may arise during meiosis, and may be referred to as a meiotic error. Unmatching trisomy may refer to the case where three copies of a given chromosome are 45 present in an individual and two of the copies are from the same parent, and are homologous, but are not identical. Note that unmatching trisomy may refer to the case where two homologous chromosomes from one parent are present, and where some segments of the chromosomes are identical 50 while other segments are merely homologous.

Homologous Chromosomes refers to chromosome copies that contain the same set of genes that normally pair up during meiosis.

Identical Chromosomes refers to chromosome copies that 55 contain the same set of genes, and for each gene they have the same set of alleles that are identical, or nearly identical.

Allele Drop Out (ADO) refers to the situation where at least one of the base pairs in a set of base pairs from homologous chromosomes at a given allele is not detected. 60

Locus Drop Out (LDO) refers to the situation where both base pairs in a set of base pairs from homologous chromosomes at a given allele are not detected.

Homozygous refers to having similar alleles as corresponding chromosomal loci.

Heterozygous refers to having dissimilar alleles as corresponding chromosomal loci.

32

Heterozygosity Rate refers to the rate of individuals in the population having heterozygous alleles at a given locus. The heterozygosity rate may also refer to the expected or measured ratio of alleles, at a given locus in an individual, or a sample of DNA or RNA.

Chromosomal Region refers to a segment of a chromosome, or a full chromosome.

Segment of a Chromosome refers to a section of a chromosome that can range in size from one base pair to the entire chromosome.

Chromosome refers to either a full chromosome, or a segment or section of a chromosome.

Copies refers to the number of copies of a chromosome segment. It may refer to identical copies, or to non-identical, homologous copies of a chromosome segment wherein the different copies of the chromosome segment contain a substantially similar set of loci, and where one or more of the alleles are different. Note that in some cases of aneuploidy, such as the M2 copy error, it is possible to have some copies of the given chromosome segment that are identical as well as some copies of the same chromosome segment that are not identical.

Haplotype refers to a combination of alleles at multiple loci that are typically inherited together on the same chromosome. Haplotype may refer to as few as two loci or to an entire chromosome depending on the number of recombination events that have occurred between a given set of loci. Haplotype can also refer to a set of SNPs on a single chromatid that are statistically associated.

Haplotypic Data, also "Phased Data" or "Ordered Genetic Data," refers to data from a single chromosome or chromosome segment in a diploid or polyploid genome, e.g., either the segregated maternal or paternal copy of a chromosome in a diploid genome.

Phasing refers to the act of determining the haplotypic genetic data of an individual given unordered, diploid (or polyploidy) genetic data. It may refer to the act of determining which of two genes at an allele, for a set of alleles found on one chromosome, are associated with each of the two homologous chromosomes in an individual.

Phased Data refers to genetic data where one or more haplotypes have been determined.

Hypothesis refers to a possible state, such as a possible degree of overrepresentation of the number of copies of a first homologous chromosome or chromosome segment as compared to a second homologous chromosome or chromosome segment, a possible deletion, a possible duplication, a possible ploidy state at a given set of one or more chromosomes or chromosome segments, a possible allelic state at a given set of one or more loci, a possible paternity relationship, or a possible DNA, RNA, fetal fraction at a given set of one or more chromosomes or chromosome segment, or a set of quantities of genetic material from a set of loci. The genetic states can optionally be linked with probabilities indicating the relative likelihood of each of the elements in the hypothesis being true in relation to other elements in the hypothesis, or the relative likelihood of the hypothesis as a whole being true. The set of possibilities may comprise one or more elements.

Copy Number Hypothesis, also "Ploidy State Hypothesis," refers to a hypothesis concerning the number of copies of a chromosome or chromosome segment in an individual. It may also refer to a hypothesis concerning the identity of each of the chromosomes, including the parent of origin of
 each chromosome, and which of the parent's two chromosomes are present in the individual. It may also refer to a hypothesis concerning which chromosomes, or chromosome

## US 11,530,454 B2

33

segments, if any, from a related individual correspond genetically to a given chromosome from an individual.

Related Individual refers to any individual who is genetically related to, and thus shares haplotype blocks with, the target individual. In one context, the related individual may be a genetic parent of the target individual, or any genetic material derived from a parent, such as a sperm, a polar body, an embryo, a fetus, or a child. It may also refer to a sibling, parent, or grandparent.

Sibling refers to any individual whose genetic parents are 10 the same as the individual in question. In some embodiments, it may refer to a born child, an embryo, or a fetus, or one or more cells originating from a born child, an embryo, or a fetus. A sibling may also refer to a haploid individual that originates from one of the parents, such as a sperm, a 15 polar body, or any other set of haplotypic genetic matter. An individual may be considered to be a sibling of itself.

Child may refer to an embryo, a blastomere, or a fetus. Note that in the presently disclosed embodiments, the concepts described apply equally well to individuals who are a 20 born child, a fetus, an embryo, or a set of cells therefrom. The use of the term child may simply be meant to connote that the individual referred to as the child is the genetic offspring of the parents.

Fetal refers to "of the fetus," or "of the region of the 25 placenta that is genetically similar to the fetus". In a pregnant woman, some portion of the placenta is genetically similar to the fetus, and the free floating fetal DNA found in maternal blood may have originated from the portion of the placenta with a genotype that matches the fetus. Note that the genetic information in half of the chromosomes in a fetus is inherited from the mother of the fetus. In some embodiments, the DNA from these maternally inherited chromosomes that came from a fetal cell is considered to be "of fetal origin," not "of maternal origin."

DNA of Fetal Origin refers to DNA that was originally part of a cell whose genotype was essentially equivalent to that of the fetus.

DNA of Maternal Origin refers to DNA that was originally part of a cell whose genotype was essentially equivalent to that of the mother.

Parent refers to the genetic mother or father of an individual. An individual typically has two parents, a mother and a father, though this may not necessarily be the case such as in genetic or chromosomal chimerism. A parent may be 45 considered to be an individual.

Parental Context refers to the genetic state of a given SNP, on each of the two relevant chromosomes for one or both of the two parents of the target.

Maternal Plasma refers to the plasma portion of the blood 50 from a female who is pregnant.

Clinical Decision refers to any decision to take or not take an action that has an outcome that affects the health or survival of an individual. A clinical decision may also refer to a decision to conduct further testing, to abort or maintain 55 a pregnancy, to take actions to mitigate an undesirable phenotype, or to take actions to prepare for a phenotype.

Diagnostic Box refers to one or a combination of machines designed to perform one or a plurality of aspects of the methods disclosed herein. In an embodiment, the 60 diagnostic box may be placed at a point of patient care. In an embodiment, the diagnostic box may perform targeted amplification followed by sequencing. In an embodiment the diagnostic box may function alone or with the help of a technician.

Informatics Based Method refers to a method that relies heavily on statistics to make sense of a large amount of data. 34

In the context of prenatal diagnosis, it refers to a method designed to determine the ploidy state at one or more chromosomes or chromosome segments, the allelic state at one or more alleles, or paternity by statistically inferring the most likely state, rather than by directly physically measuring the state, given a large amount of genetic data, for example from a molecular array or sequencing. In an embodiment of the present disclosure, the informatics based technique may be one disclosed in this patent application. In an embodiment of the present disclosure it may be PARENTAL SUPPORT.

Primary Genetic Data refers to the analog intensity signals that are output by a genotyping platform. In the context of SNP arrays, primary genetic data refers to the intensity signals before any genotype calling has been done. In the context of sequencing, primary genetic data refers to the analog measurements, analogous to the chromatogram, that comes off the sequencer before the identity of any base pairs have been determined, and before the sequence has been mapped to the genome.

Secondary Genetic Data refers to processed genetic data that are output by a genotyping platform. In the context of a SNP array, the secondary genetic data refers to the allele calls made by software associated with the SNP array reader, wherein the software has made a call whether a given allele is present or not present in the sample. In the context of sequencing, the secondary genetic data refers to the base pair identities of the sequences have been determined, and possibly also where the sequences have been mapped to the genome.

Preferential Enrichment of DNA that corresponds to a locus, or preferential enrichment of DNA at a locus, refers to any method that results in the percentage of molecules of DNA in a post-enrichment DNA mixture that correspond to the locus being higher than the percentage of molecules of DNA in the pre-enrichment DNA mixture that correspond to the locus. The method may involve selective amplification of DNA molecules that correspond to a locus. The method may involve removing DNA molecules that do not correspond to the locus. The method may involve a combination of methods. The degree of enrichment is defined as the percentage of molecules of DNA in the post-enrichment mixture that correspond to the locus divided by the percentage of molecules of DNA in the pre-enrichment mixture that correspond to the locus. Preferential enrichment may be carried out at a plurality of loci. In some embodiments of the present disclosure, the degree of enrichment is greater than 20, 200, or 2,000. When preferential enrichment is carried out at a plurality of loci, the degree of enrichment may refer to the average degree of enrichment of all of the loci in the

Amplification refers to a method that increases the number of copies of a molecule of DNA or RNA.

Selective Amplification may refer to a method that increases the number of copies of a particular molecule of DNA (or RNA), or molecules of DNA (or RNA) that correspond to a particular region of DNA (or RNA). It may also refer to a method that increases the number of copies of a particular targeted molecule of DNA (or RNA), or targeted region of DNA (or RNA) more than it increases non-targeted molecules or regions of DNA (or RNA). Selective amplification may be a method of preferential enrichment.

Universal Priming Sequence refers to a DNA (or RNA) sequence that may be appended to a population of target DNA (or RNA) molecules, for example by ligation, PCR, or ligation mediated PCR. Once added to the population of target molecules, primers specific to the universal priming

## US 11,530,454 B2

3:

sequences can be used to amplify the target population using a single pair of amplification primers. Universal priming sequences are typically not related to the target sequences.

Universal Adapters, or "ligation adaptors" or "library tags" are nucleic acid molecules containing a universal 5 priming sequence that can be covalently linked to the 5-prime and 3-prime end of a population of target double stranded nucleic acid molecules. The addition of the adapters provides universal priming sequences to the 5-prime and 3-prime end of the target population from which PCR 10 amplification can take place, amplifying all molecules from the target population, using a single pair of amplification primers.

Targeting refers to a method used to selectively amplify or otherwise preferentially enrich those molecules of DNA (or 15 RNA) that correspond to a set of loci in a mixture of DNA (or RNA)

Joint Distribution Model refers to a model that defines the probability of events defined in terms of multiple random variables, given a plurality of random variables defined on 20 the same probability space, where the probabilities of the variable are linked. In some embodiments, the degenerate case where the probabilities of the variables are not linked may be used.

Cancer-related gene refers to a gene associated with an 25 altered risk for a cancer or an altered prognosis for a cancer. Exemplary cancer-related genes that promote cancer include oncogenes; genes that enhance cell proliferation, invasion, or metastasis; genes that inhibit apoptosis; and pro-angiogenesis genes. Cancer-related genes that inhibit cancer include, but are not limited to, tumor suppressor genes; genes that inhibit cell proliferation, invasion, or metastasis; genes that promote apoptosis; and anti-angiogenesis genes.

Estrogen-related cancer refers to a cancer that is modulated by estrogen. Examples of estrogen-related cancers 3 include, without limitation, breast cancer and ovarian cancer. Her2 is overexpressed in many estrogen-related cancers (U.S. Pat. No. 6,165,464, which is hereby incorporated by reference in its entirety).

Androgen-related cancer refers to a cancer that is modulated by androgen. An example of androgen-related cancers is prostate cancer.

Higher than normal expression level refers to expression of an mRNA or protein at a level that is higher than the average expression level of the corresponding molecule in 45 control subjects (such as subjects without a disease or disorder such as cancer). In various embodiments, the expression level is at least 20, 40, 50, 75, 90, 100, 200, 500, or even 1000% higher than the level in control subjects.

Lower than normal expression level refers to expression 50 of an mRNA or protein at a level that is lower than the average expression level of the corresponding molecule in control subjects (such as subjects without a disease or disorder such as cancer). In various embodiments, the expression level is at least 20, 40, 50, 75, 90, 95, or 100% 55 lower than the level in control subjects. In some embodiments, the expression of the mRNA or protein is not detectable.

Modulate expression or activity refers to either increasing or decreasing expression or activity, for example, of a 60 protein or nucleic acid sequence, relative to control conditions. In some embodiments, the modulation in expression or activity is an increase or decrease of at least 10, 20, 40, 50, 75, 90, 100, 200, 500, or even 1000%. In various embodiments, transcription, translation, mRNA or protein 65 stability, or the binding of the mRNA or protein to other molecules in vivo is modulated by the therapy. In some

36

embodiments, the level of mRNA is determined by standard Northern blot analysis, and the level of protein is determined by standard Western blot analysis, such as the analyses described herein or those described by, for example, Ausubel et al. (Current Protocols in Molecular Biology, John Wiley & Sons, New York, Jul. 11, 2013, which is hereby incorporated by reference in its entirety). In one embodiment, the level of a protein is determined by measuring the level of enzymatic activity, using standard methods. In another preferred embodiment, the level of mRNA, protein, or enzymatic activity is equal to or less than 20, 10, 5, or 2-fold above the corresponding level in control cells that do not express a functional form of the protein, such as cells homozygous for a nonsense mutation. In yet another embodiment, the level of mRNA, protein, or enzymatic activity is equal to or less than 20, 10, 5, or 2-fold above the corresponding basal level in control cells, such as noncancerous cells, cells that have not been exposed to conditions that induce abnormal cell proliferation or that inhibit apoptosis, or cells from a subject without the disease or disorder of interest.

Dosage sufficient to modulate mRNA or protein expression or activity refers to an amount of a therapy that increases or decreases mRNA or protein expression or activity when administered to a subject. In some embodiments, for a compound that decreases expression or activity, the modulation is a decrease in expression or activity that is at least 10%, 30%, 40%, 50%, 75%, or 90% lower in a treated subject than in the same subject prior to the administration of the inhibitor or than in an untreated, control subject. In addition, In some embodiments, for a compound that increases expression or activity, the amount of expression or activity of the mRNA or protein is at least 1.5-, 2-, 3-, 5-, 10-, or 20-fold greater in a treated subject than in the same subject prior to the administration of the modulator or than in an untreated, control subject.

In some embodiments, compounds may directly or indirectly modulate the expression or activity of the mRNA or protein. For example, a compound may indirectly modulate the expression or activity of an mRNA or protein of interest by modulating the expression or activity of a molecule (e.g., a nucleic acid, protein, signaling molecule, growth factor, cytokine, or chemokine) that directly or indirectly affects the expression or activity of the mRNA or protein of interest. In some embodiments, the compounds inhibit cell division or induce apoptosis. These compounds in the therapy may include, for example, unpurified or purified proteins, antibodies, synthetic organic molecules, naturally-occurring organic molecules, nucleic acid molecules, and components thereof. The compounds in a combination therapy may be administered simultaneously or sequentially. Exemplary compounds include signal transduction inhibitors.

Purified refers to being separated from other components that naturally accompany it. Typically, a factor is substantially pure when it is at least 50%, by weight, free from proteins, antibodies, and naturally-occurring organic molecules with which it is naturally associated. In some embodiments, the factor is at least 75%, 90%, or 99%, by weight, pure. A substantially pure factor may be obtained by chemical synthesis, separation of the factor from natural sources, or production of the factor in a recombinant host cell that does not naturally produce the factor. Proteins and small molecules may be purified by one skilled in the art using standard techniques such as those described by Ausubel et al. (Current Protocols in Molecular Biology, John Wiley & Sons, New York, Jul. 11, 2013, which is hereby incorporated by reference in its entirety). In some embodiments the factor

## US 11,530,454 B2

3'

is at least 2, 5, or 10 times as pure as the starting material, as measured using polyacrylamide gel electrophoresis, column chromatography, optical density, HPLC analysis, or western analysis (Ausubel et al., supra). Exemplary methods of purification include immunoprecipitation, column chromatography such as immunoaffinity chromatography, magnetic bead immunoaffinity purification, and panning with a plate-bound antibody.

Other features and advantages of the invention will be apparent from the following detailed description and from the claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

The presently disclosed embodiments will be further explained with reference to the attached drawings, wherein like structures are referred to by like numerals throughout the several views. The drawings shown are not necessarily to scale, with emphasis instead generally being placed upon illustrating the principles of the presently disclosed embodi- 25 ments.

FIGS. 1A-1D are graphs showing the distribution of the test statistic S divided by T (the number of SNPs) ("S/T") for various copy number hypotheses for a depth of read (DOR) of 500 and a tumor fraction of 1% for an increasing number of SNPs. FIG. 1A: 100 SNPs, FIG. 1B: 333 SNPs, FIG. 1C: 667 SNPs, FIG. 1D: 1000 SNPs.

FIGS. **2**A-**2**D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 500 and tumor fraction of 2% for an increasing number of SNPs. 35 FIG. **2**A: 100 SNPs, FIG. **2**B: 333 SNPs, FIG. **2**C: 667 SNPs, FIG. **2**D: 1000 SNPs.

FIGS. **3A-3**D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 500 and tumor fraction of 3% for an increasing number of SNPs. 40 FIG. **3A**: 100 SNPs, FIG. **3B**: 333 SNPs, FIG. **3C**: 667 SNPs, FIG. **3D**: 1000 SNPs.

FIGS. 4A-4D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 500 and tumor fraction of 4% for an increasing number of SNPs. 45 FIG. 4A: 100 SNPs, FIG. 4B: 333 SNPs, FIG. 4C: 667 SNPs, FIG. 4D: 1000 SNPs.

FIGS. 5A-5D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 500 and tumor fraction of 5% for an increasing number of SNPs. 50 FIG. 5A: 100 SNPs, FIG. 5B: 333 SNPs, FIG. 5C: 667 SNPs, FIG. 5D: 1000 SNPs.

FIGS. **6A-6D** are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 500 and tumor fraction of 6% for an increasing number of SNPs. 55 FIG. **6A**: 100 SNPs, FIG. **6B**: 333 SNPs, FIG. **6C**: 667 SNPs, FIG. **6D**: 1000 SNPs.

FIGS. 7A-7D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 1000 and tumor fraction of 0.5% for an increasing number of SNPs. 60 FIG. 7A: 100 SNPs, FIG. 7B: 333 SNPs, FIG. 7C: 667 SNPs, FIG. 7D: 1000 SNPs.

FIGS. **8A-8D** are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 1000 and tumor fraction of 1% for an increasing number of SNPs. 65 FIG. **8A**: 100 SNPs, FIG. **8B**: 333 SNPs, FIG. **8C**: 667 SNPs, FIG. **8D**: 1000 SNPs.

38

FIGS. **9A-9**D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 1000 and tumor fraction of 2% for an increasing number of SNPs. FIG. **9A**: 100 SNPs, FIG. **9B**: 333 SNPs, FIG. **9C**: 667 SNPs, FIG. **9D**: 1000 SNPs.

FIGS. 10A-10D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 1000 and tumor fraction of 3% for an increasing number of SNPs. FIG. 10A: 100 SNPs, FIG. 10B: 333 SNPs, FIG. 10C: 667 SNPs, FIG. 10D: 1000 SNPs.

FIGS. 11A-11D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 1000 and tumor fraction of 4% for an increasing number of SNPs. FIG. 11A: 100 SNPs, FIG. 11B: 333 SNPs, FIG. 11C: 667 SNPs, FIG. 11D: 1000 SNPs.

FIGS. **12**A-**12**D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 3000 and tumor fraction of 0.5% for an increasing number of SNPs. FIG. **12**A: 100 SNPs, FIG. **12**B: 333 SNPs, FIG. **12**C: 667 SNPs, FIG. **12**D: 1000 SNPs.

FIGS. **13**A-**13**D are graphs showing the distribution of S/T for various copy number hypotheses for a DOR of 3000 and tumor fraction of 1% for an increasing number of SNPs. FIG. **13**A: 100 SNPs, FIG. **13**B: 333 SNPs, FIG. **13**C: 667 SNPs, FIG. **13**D: 1000 SNPs.

FIG. 14 is a table indicating the sensitivity and specificity for detecting six microdeletion syndromes.

FIG. 15 is a graphical representation of euploidy. The x-axis represents the linear position of the individual polymorphic loci along the chromosome, and the y-axis represents the number of A allele reads as a fraction of the total (A+B) allele reads. Maternal and fetal genotypes are indicated to the right of the plots. The plots are symbol-coded according to maternal genotype, such that solid circles indicate a maternal genotype of AA, solid squares indicate a maternal genotype of BB, and open triangles indicate a maternal genotype of AB. The left plot is a plot of when two chromosomes are present, and the fetal cfDNA fraction is 0%. This plot is from a non-pregnant woman, and thus represents the pattern when the genotype is entirely maternal. Allele clusters are thus centered around 1 (AA alleles), 0.5 (AB alleles), and 0 (BB alleles). The center plot is a plot of when two chromosomes are present, and the fetal fraction is 12%. The contribution of fetal alleles to the fraction of A allele reads shifts the position of some allele spots up or down along the y-axis. The right plot is a plot of when two chromosomes are present, and the fetal fraction is 26%. The pattern, including two solid circle and two solid square peripheral bands and a trio of central open triangles, is readily apparent.

FIGS. 16A and 16B are graphical representations of 22q11.2 deletion syndrome. FIG. 16A is for maternal 22q11.2 deletion carrier (as indicated by the absence of the open triangles indicating AB SNPs). FIG. 16B is for a paternally inherited 22q11 deletion in a fetus (as indicated by the presence of solid circle and solid square peripheral bands). The x-axis represents the linear position of the SNPs, and the y-axis indicates the fraction of A allele reads out of the total reads. Each individual circle, triangle or square represents a single SNP locus.

FIG. 17 is a graphical representation of maternally inherited Cri-du-Chat deletion syndrome (as indicated by the presence of two central open triangle shape bands instead of three open triangle shape bands). The x-axis represents the linear position of the SNPs, and the y-axis indicates the

fraction of A allele reads out of the total reads. Each individual circle, triangle or square represents a single SNP

FIG. 18 is a graphical representation of paternally inherited Wolf-Hirschhorn deletion syndrome (as indicated by the presence of solid circle and solid square peripheral bands). The x-axis represents the linear position of the SNPs, and the y-axis indicates the fraction of A allele reads out of the total reads. Each individual circle, triangle or square represents a single SNP locus.

FIGS. 19A-19D are graphical representations of X chromosome spike-in experiments to represent an extra copy of a chromosome or chromosome segment. The plots show different amounts of DNA from a father mixed with DNA from the daughter: 16% father DNA (FIG. 19A), 10% father 15 DNA (FIG. 19B), 1% father DNA (FIG. 19C), and 0.1% father DNA (FIG. 19D). The x-axis represents the linear position of the SNPs on the X chromosome, and the y-axis indicates the fraction of M allele reads out of the total reads (M+R). Each individual criss-cross, circle, triangle or square 20 represents a single SNP locus with allele M or R

FIGS. 20A and 20B are graphs of the false negative rate using haplotype data (FIG. 20A) and without haplotype data

FIGS. 21A and 21B are graphs of the false positive rate 25 for p=1% using haplotype data (FIG. 21A) and without haplotype data (FIG. 21B).

FIGS. 22A and 22B are graphs of the false positive rate for p=1.5% using haplotype data (FIG. 22A) and without haplotype data (FIG. 22B).

FIGS. 23A and 23B are graphs of the false positive rate for p=2% using haplotype data (FIG. 23A) and without haplotype data (FIG. 23B).

FIGS. 24A and 24B are graphs of the false positive rate for p=2.5% using haplotype data (FIG. 24A) and without 35 haplotype data (FIG. 24B).

FIGS. 25A and 25B are graphs of the false positive rate for p=3% using haplotype data (FIG. 25A) and without haplotype data (FIG. 25B).

FIG. 27 is a table of false negative rates for the first simulation.

FIG. 28 contains a graph of reference counts (counts of one allele, such as the "A" allele) divided by total counts for 45 that locus for a normal (noncancerous) cell line, a graph of reference counts divided by total counts for a cancer cell line with a deletion and a graph of reference counts divided by total counts for a mixture of DNA from the normal cell line (95%) and the cancer cell line (5%).

FIG. 29 is a graph of reference counts divided by total counts for a plasma sample from a patient with stage IIa breast cancer with a tumor fraction estimated to be 4.33% (in which 4.33% of the DNA is from tumor cells). The diamond portion of the graph represents a region in which no CNV is 55 present. The portion of the graph with solid circles and squares represents a region in which a CNV is present and there is a visible separation of the measured allele ratios from the expected allele ratio of 0.5. The solid square indicates one haplotype, and the solid circle indicates the 60 other haplotype. Approximately 636 heterozygous SNPs were analyzed in the region of the CNV.

FIG. 30 is a graph of reference counts divided by total counts for a plasma sample from a patient with stage IIb breast cancer with a tumor fraction estimated to be 0.58%. 65 The open diamonds of the graph represents a region in which no CNV is present. The portion of the graph with solid

circles and squares represents a region in which a CNV is present but there is no clearly visible separation of the measured allele ratios from the expected allele ratio of 0.5. For this analysis, 86 heterozygous SNPs were analyzed in the region of the CNV.

FIGS. 31A and 31B are graphs showing the maximum likelihood estimation of the tumor fraction. The maximum likelihood estimate is indicated by the peak of the graph and is 4.33% for FIG. 31A and 0.58% for FIG. 31B.

FIG. 32A is a comparison of the graphs of the log of the odds ratio for various possible tumor fractions for the high tumor fraction sample (4.33%) and the low tumor fraction sample (0.58%). If the log odds ratio is less than 0, the euploid hypothesis is more likely. If the log odds ratio is greater than 0, the presence of a CNV is more likely.

FIG. 32B is a graph of small tumor results plotted in probability space. The graph depicts the probability of a deletion divided by the probability of no deletion for various possible tumor fractions for the low tumor fraction sample (0.58%).

FIG. 33 is a graph of the log of the odds ratio for various possible tumor fractions for the low tumor fraction sample (0.58%). FIG. 33 is an enlarged version of the graph in FIG. 32A for the low tumor fraction sample.

FIG. 34 is a graph showing the limit of detection for single nucleotide variants in a tumor biopsy using three different methods described in Example 6.

FIG. 35 is a graph showing the limit of detection for single nucleotide variants in a plasma sample using three different 30 methods described in Example 6.

FIGS. 36A and 36B are graphs of the analysis of genomic DNA (FIG. 36A) or DNA from a single cell (FIG. 36B) using a library of approximately 28,000 primers designed to detect CNVs. The presence of two central bands instead of one central band indicates the presence of a CNV. The x-axis represents the linear position of the SNPs, and the y-axis indicates the fraction of A allele reads out of the total reads.

FIGS. 37A and 37B are graphs of the analysis of genomic DNA (FIG. 37A) or DNA from a single cell (FIG. 37B) FIG. 26 is a table of false positive rates for the first 40 using a library of approximately 3,000 primers designed to detect CNVs. The presence of two central bands instead of one central band indicates the presence of a CNV. The x-axis represents the linear position of the SNPs, and the y-axis indicates the fraction of A allele reads out of the total reads.

FIG. 38 is a graph illustrating the uniformity in DOR for these ~3.000 loci.

FIG. 39 is a table comparing error call metrics for genomic DNA and DNA from a single cell.

FIG. 40 is a graph of error rates for transition mutations and transversion mutations.

FIGS. 41A-D are graphs of Sensitivity of CoNVERGe determined with PlasmArts. FIG. 41A: Correlation between CoNVERGe-calculated AAI and actual input fraction in PlasmArt samples with DNA from a 22q11.2 deletion and matched normal cell lines. FIG. 41B: Correlation between calculated AAI and actual tumour DNA input in PlasmArt samples with DNA from HCC2218 breast cancer cells with chromosome 2p and 2q CNVs and matched normal HCC2218BL cells, containing 0-9.09% tumour DNA fractions. FIG. 41C: Correlation between calculated AAI and actual tumour DNA input in PlasmArt samples with DNA from HCC1954 breast cancer cells with chromosome 1p and 1q CNVs and matched normal HCC1954BL cells, containing 0-5.66% tumour DNA fractions. FIG. 41D: Allele frequency plot for HCC1954 cells used in FIG. C. In FIGS. 41A-C, data points and error bars indicate the mean and standard deviation (SD), respectively, of 3-8 replicates.

FIGS. 42A-B provide a model system for validation. Plasmart samples were made from cell lines with similar size profiles to plasma. FIG. 42A illustrates a son's plasma with a 22q11 deletion spiked into the father's plasma. Focal CNV: 3 1\4B. FIG. 42B illustrates Chromosomes 1 and 2: cancer cell lines into normal cell line of same individual. CNVs on chromosome arms 1p, 1q, 2p, 2q. FIGS. 42A and 42B are graphs showing fragment size distributions of an exemplary Plasmart standard.

FIG. 43A, FIG. 43B, FIG. 43C, and FIG. 43D provide results from a dilution curve of Plasmart synthetic ctDNA standards for validation of microdeletion and cancer panels. FIG. 43A is a graph showing the maximum likelihood of tumor. FIG. 43B is an estimate of DNA fraction results as an 15 odds ratio plot. FIG. 43C is a plot for the detection of transversion events. FIG. 43D is a plot for the detection of Transition events.

FIG. 44 is a plot showing CNVs for various chromosomal ctDNAs. The plot depicts plasma from 21 breast cancer patients (stage 1-IIIB) and demonstrated that CNVs could be detected in ctDNA with an AAI≥0.45% and required as few as 62 heterozygous SNPs.

FIG. 45 is a plot showing CNVs for various chromosomal 25 regions for various ovarian cancer samples with different % ctDNA levels. The plot indicates 100% detection rate at a 9.45% cutoff.

FIG. 46A is a table showing the percent of breast or lung cancer patients with an SNV or a combined SNV and/or 30 CNV in ctDNA. The analysis was on ctDNA (plasma) from Stage I-III cancer patients and indicates that the ability to detect CNV in plasma dramatically improves detection rate vs. testing SNVs alone. FIG. 46B plots cumulative proportion TCGA breast cancer patients covered vs. genes with breast SNVs. FIG. 46C plots cumulative COSMIC patient capture vs. cumulative patient coverage (TCGA) for breast deletions. FIG. 46D plots cumulative COSMIC patient capture vs. cumulative patient coverage (TCGA) for breast amplifications.

FIG. 47A is a graph of % samples at different breast cancer stages with tumor-specific SNVs and/or CNVs in plasma. FIG. 47B is a table of percent detection of breast CNVs and SNVs by stage.

FIG. 48A is a graph of % samples at different breast 45 cancer substages with tumor-specific SNVs and/or CNVs in plasma. FIG. 48B is a table of percent detection of breast CNVs and SNVs by tumor substage.

FIG.  $49\mathrm{A}\xspace$  is a graph of % samples at different lung cancer stages with tumor-specific SNVs and/or CNVs in plasma. 50 FIG. 49B is a table of lung plasma detection rate of lung SNVs and/or CNVs.

FIG. 50A is a graph of % samples at different lung cancer substages with tumor-specific SNVs and/or CNVs in plasma. FIG. 50B is a table of lung plasma detection rate of 55 lung SNVs and/or CNVs by tumor substage.

FIG. 51A represents the histological finding/history for primary lung tumors analyzed for clonal and subclonal tumor heterogeneity. FIG. 51B is a table of the VAF identities of the biopsied lung tumors by whole genome sequenc- 60 ing and assaying by AmpliSEQ.

FIG. 52 illustrates the use of ctDNA from plasma to identify both clonal and subclonal SNV mutations to overcome tumor heterogeneity.

FIG. 53A is a table comparing VAF calls by AmpliSeq. 65 FIG. 53B is a table comparing VAF calls by mmPCR-NGS. A comparison of the two tables for detection of SNVs in

primary tumor indicate that SNVs were missed by AmpliSeq and SNV mutations were identified in ctDNA from plasma with mmPCR-NGS

FIG. 54A is a plot of % VAF in Primary Lung Tumor. FIG. 54B is a linear regression plot of AmpliSeq VAF vs. Natera VAF

FIG. 55 is a graph of Pool 1/4 of an 84-plex SNV PCR primer reaction when primer concentration is limited.

FIG. 56 is a graph of Pool 2/4 of an 84-plex SNV PCR rimer reaction when primer concentration is limited.

FIG. 57 is a graph of Pool 3/4 of an 84-plex SNV PCR primer reaction when primer concentration is limited.

FIG. 58 is a graph of Pool 4/4 of an 84-plex SNV PCR primer reaction when primer concentration is limited.

FIG. 59 illustrates a plot of Limit of Detection (LOD) vs. Depth of Read (DOR) for detection of SNV Transition and Transversion mutations in a 84-plex PCR reaction at 15 PCR cycles.

FIG. 60 illustrates a plot of Limit of Detection (LOD) vs. regions as indicated for various samples at different % 20 Depth of Read (DOR) for detection of SNV Transition and Transversion mutations in a 84-plex PCR reaction at 20 PCR

> FIG. 61 illustrates a plot of Limit of Detection (LOD) vs. Depth of Read (DOR) for detection of SNV Transition and Transversion mutations in a 84-plex PCR reaction at 25 PCR

> FIG. 62A is a plot illustrating sensitivity of detection of SNVs in tumor cell genomic DNA. FIG. 62B illustrates sensitivity of detection of SNVs in 1/3 single cells. FIG. 62C illustrates sensitivity of detection of SNVs in 2/3 single cells. FIG. 62D illustrates sensitivity of detection of SNVs in 3/3 single cells. Comparable sensitivities are seen between tumor and single cell genomic DNA.

> FIG. 63A illustrates the workflow for analysis of CNVs in a variety of cancer sample types in a massively multiplexed PCR (mmPCR) assay targeting SNPs. FIG. 63B illustrates detection of somatic CNVs in human breast cancer cell lines and matched normal cell lines (FIG. 63C) on the CoN-VERGe platform. FIG. 63D illustrates detection of somatic CNVs in human breast cancer cell lines and matched normal cell lines (FIG. 63E) on the CytoSNP-12 microarray platform. FIG. 63F is a plot of the maximum homolog ratios for CNVs identified by CoNVERG3e or CytoSNP-12 showing a strong linear correlation of identified CNVs by either method.

> FIGS. 64A-H provide a comparison of Fresh Frozen (FF) and FFPE (formalin-fixed paraffin embedded) breast cancer samples to matched buffy coat gDNA control samples. FIG. 64A is a FF breast tissue control sample analyzed by CoNVERGe. FIG. 64B is a FFPE breast tissue control sample analyzed by CoNVERGe. FIG. 64C is a FF breast tumour tissue sample analyzed by CoNVERGe. FIG. 64D is a FFPE breast tumour tissue sample analyzed by CoN-VERGe. FIG. 64E is a FF breast tumour tissue sample analyzed by CytoSNP-12. FIG. 64F is a FFPE breast tumour tissue sample analyzed by CytoSNP-12. FIG.  $\bf 64$ G compares the CoNVERGe assay to a microarray assay on breast cancer cell lines and FIG. 64H compares the CoNVERGe assay to the OneScan assay on breast cancer cell lines.

> FIGS. 65A-D illustrate Allele frequency plots to reflect chromosome copy number using the CoNVERGe assay to detect CNVs in single cells. FIG. 65A is the analysis of 1/3 breast cancer single cell replicates. FIG.  ${\bf 65}{\rm B}$  is the analysis of 2/3 breast cancer single cell replicates. FIG. 65C is the analysis of 3/3 breast cancer single cell replicates. FIG. 65D is the analysis of a B-lymphocyte cell line lacking CNVs in the target regions.

# US 11,530,454 B2

4

FIGS. 66A-C illustrate Allele frequency plots to reflect chromosome copy number using the CoNVERGe assay to detect CNVs in real plasma samples. FIG. 66A is a stage II breast cancer plasma cfDNA sample and its matched tumor biopsy gDNA. FIG. 66B is a late stage ovarian cancer plasma cfDNA sample and its matched tumor biopsy gDNA. FIG. 66C is a chart illustrating tumor heterogeneity as determined by CNV detection in five late stage ovarian cancer plasma and matched tissue samples.

FIGS. 67A-H lists the chromosome positions, SNVs and 10 mutation change in breast cancer.

FIGS. **68**A-B illustrate the major (FIG. **68**A) and minor allele (FIG. **68**B) frequencies of SNPs used in a 3168 mmPCR reaction.

FIG. 69 shows an example system architecture X00 useful 15 for performing embodiments of the present invention. System architecture X00 includes an analysis platform X08 and a laboratory information systems ("LISs") X04. X04 can be connected to Genetic Data Source X10. X08 may be connected to LIS X04 over a network X02. Analysis platform X08 may alternatively or additionally be connected directly to LIS X06. LIS X06 can be connected to Genetic Data Source X10. Analysis platform X08 includes one or more of an input processor X12, a hypothesis manager X14, a modeler X16, an error correction unit X18, a machine 25 learning unit X20, and an output processor X22.

FIG. 70 illustrates an example computer system Y00 for performing embodiments of the present invention. System architecture Y00 includes one or more processors Y10, a BUS Y20, a main memory Y30, a memory controller Y75, a communications and network interface Y80, a communication path Y85, an input/output/display devices Y90, and may also include a secondary memory Y40. Y40 may include a hard disk drive Y50 and a removable storage drive Y60. Y60 can write to a removable storage unit Y70.

While the above-identified drawings set forth presently disclosed embodiments, other embodiments are also contemplated, as noted in the discussion. This disclosure presents illustrative embodiments by way of representation and not limitation. Numerous other modifications and embodiments can be devised by those skilled in the art which fall within the scope and spirit of the principles of the presently disclosed embodiments.

# DETAILED DESCRIPTION OF THE INVENTION

In one aspect, the present invention generally relates, at least in part, to improved methods of determining the presence or absence of copy number variations, such as 50 deletions or duplications of chromosome segments or entire chromosomes. The methods are particularly useful for detecting small deletions or duplications, which can be difficult to detect with high specificity and sensitivity using prior methods due to the small amount of data available from 55 the relevant chromosome segment. The methods include improved analytical methods, improved bioassay methods, and combinations of improved analytical and bioassay methods. Methods of the invention can also be used to detect deletions or duplications that are only present in a small 60 percentage of the cells or nucleic acid molecules that are tested. This allows deletions or duplications to be detected prior to the occurrence of disease (such as at a precancerous stage) or in the early stages of disease, such as before a large number of diseased cells (such as cancer cells) with the 65 deletion or duplication accumulate. The more accurate detection of deletions or duplications associated with a

44

disease or disorder enable improved methods for diagnosing, prognosticating, preventing, delaying, stabilizing, or treating the disease or disorder. Several deletions or duplications are known to be associated with cancer or with severe mental or physical handicaps.

In another aspect, the present invention generally relates, at least in part, to improved methods of detecting single nucleotide variations (SNVs). These improved methods include improved analytical methods, improved bioassay methods, and improved methods that use a combination of improved analytical and bioassay methods. The methods in certain illustrative embodiments are used to detect, diagnose, monitor, or stage cancer, for example in samples where the SNV is present at very low concentrations, for example less than 10%, 5%, 4%, 3%, 2.5%, 2%, 1%, 0.5%, 0.25%, or 0.1% relative to the total number of normal copies of the SNV locus, such as circulating free DNA samples. That is, these methods in certain illustrative embodiments are particularly well suited for samples where there is a relatively low percentage of a mutation or variant relative to the normal polymorphic alleles present for that genetic loci. Finally, provided herein are methods that combine the improved methods for detecting copy number variations with the improved methods for detecting single nucleotide variations.

Successful treatment of a disease such as cancer often relies on early diagnosis, correct staging of the disease, selection of an effective therapeutic regimen, and close monitoring to prevent or detect relapse. For cancer diagnosis, histological evaluation of tumor material obtained from tissue biopsy is often considered the most reliable method. However, the invasive nature of biopsy-based sampling has rendered it impractical for mass screening and regular follow up. Therefore, the present methods have the advantage of being able to be performed non-invasively if desired for relatively low cost with fast turnaround time. The targeted sequencing that may be used by the methods of the invention requires less reads than shotgun sequencing, such as a few million reads instead of 40 million reads, thereby decreasing cost. The multiplex PCR and next generation sequencing that may be used increase throughput and reduces costs.

In some embodiments, the methods are used to detect a deletion, duplication, or single nucleotide variant in an 45 individual. A sample from the individual that contains cells or nucleic acids suspected of having a deletion, duplication, or single nucleotide variant may be analyzed. In some embodiments, the sample is from a tissue or organ suspected of having a deletion, duplication, or single nucleotide variant, such as cells or a mass suspected of being cancerous. The methods of the invention can be used to detect deletion, duplication, or single nucleotide variant that are only present in one cell or a small number of cells in a mixture containing cells with the deletion, duplication, or single nucleotide variant and cells without the deletion, duplication, or single nucleotide variant. In some embodiments, cfDNA or cfRNA from a blood sample from the individual is analyzed. In some embodiments, cfDNA or cfRNA is secreted by cells, such as cancer cells. In some embodiments, cfDNA or cfRNA is released by cells undergoing necrosis or apoptosis, such as cancer cells. The methods of the invention can be used to detect deletion, duplication, or single nucleotide variant that are only present in a small percentage of the cfDNA or cfRNA. In some embodiments, one or more cells from an embryo are tested.

In some embodiments, the methods are used for non-invasive or invasive prenatal testing of a fetus. These

# US 11,530,454 B2

4

methods can be used to determine the presence or absence of deletions or duplications of a chromosome segment or an entire chromosome, such as deletions or duplications known to be associated severe mental or physical handicaps, learning disabilities, or cancer. In some embodiments for noninvasive prenatal testing (NIPT), cells, cfDNA or cfRNA from a blood sample from the pregnant mother is tested. The methods allow the detection of a deletion or duplication in the cells, cfDNA, or cfRNA from the fetus despite the large amount of cells, cfDNA, or cfRNA from the mother that is 10 also present. In some embodiments for invasive prenatal testing, DNA or RNA from a sample from the fetus is tested (such as a CVS or amniocentesis sample). Even if the sample is contaminated with DNA or RNA from the pregnant mother, the methods can be used to detect a deletion or 15 duplication in the fetal DNA or RNA.

In addition to determining the presence or absence of copy number variation, one or more other factors can be analyzed if desired. These factors can be used to increase the accuracy of the diagnosis (such as determining the presence 20 or absence of cancer or an increased risk for cancer, classifying the cancer, or staging the cancer) or prognosis. These factors can also be used to select a particular therapy or treatment regimen that is likely to be effective in the subject. Exemplary factors include the presence or absence of polymorphisms or mutation; altered (increased or decreased) levels of total or particular cfDNA, cfRNA, microRNA (miRNA); altered (increased or decreased) tumor fraction; altered (increased or decreased) DNA integrity, altered (increased or decreased) or alternative mRNA splicing.

The following sections describe methods for detecting deletions or duplications using phased data (such as inferred or measured phased data) or unphased data; samples that can be tested; methods for sample preparation, amplification, and quantification; methods for phasing genetic data; polymorphisms, mutations, nucleic acid alterations, mRNA splicing alterations, and changes in nucleic acid levels that can be detected; databases with results from the methods, other risk factors and screening methods; cancers that can be diagnosed or treated; cancer treatments; cancer models for testing treatments; and methods for formulating and administering treatments.

Exemplary Methods for Determining Ploidy Using Phased Data

Some of the methods of the invention are based in part on the discovery that using phased data for detecting CNVs decreases the false negative and false positive rates compared to using unphased data (FIGS. 20A-27). This improvement is greatest for samples with CNVs present in 50 low levels. Thus, phase data increases the accuracy of CNV detection compared to using unphased data (such as methods that calculate allele ratios at one or more loci or aggregate allele ratios to give an aggregated value (such as an average value) over a chromosome or chromosome segment without 55 considering whether the allele ratios at different loci indicate that the same or different haplotypes appear to be present in an abnormal amount). Using phased data allows a more accurate determination to be made of whether differences between measured and expected allele ratios are due to noise 60 or due to the presence of a CNV. For example, if the differences between measured and expected allele ratios at most or all of the loci in a region indicate that the same haplotype is overrepresented, then a CNV is more likely to be present. Using linkage between alleles in a haplotype 65 allows one to determine whether the measured genetic data is consistent with the same haplotype being overrepresented

46

(rather than random noise). In contrast, if the differences between measured and expected allele ratios are only due to noise (such as experimental error), then in some embodiments, about half the time the first haplotype appears to be overrepresented and about the other half of the time, the second haplotype appears to be overrepresented.

Accuracy can be increased by taking into account the linkage between SNPs, and the likelihood of crossovers having occurred during the meiosis that gave rise to the gametes that formed the embryo that grew into the fetus. Using linkage when creating the expected distribution of allele measurements for one or more hypotheses allows the creation of expected allele measurements distributions that correspond to reality considerably better than when linkage is not used. For example, imagine that there are two SNPs, 1 and 2 located nearby one another, and the mother is A at SNP 1 and A at SNP 2 on one homolog, and B at SNP 1 and B at SNP 2 on homolog two. If the father is A for both SNPs on both homologs, and a B is measured for the fetus SNP 1, this indicates that homolog two has been inherited by the fetus, and therefore that there is a much higher likelihood of a B being present in the fetus at SNP 2.  $\widetilde{A}$  model that takes into account linkage can predict this, while a model that does not take linkage into account cannot. Alternately, if a mother is AB at SNP 1 and AB at nearby SNP 2, then two hypotheses corresponding to maternal trisomy at that location can be used—one involving a matching copy error (nondisjunction in meiosis II or mitosis in early fetal development), and one involving an unmatching copy error (nondisjunction in meiosis I). In the case of a matching copy error trisomy, if the fetus inherited an AA from the mother at SNP 1, then the fetus is much more likely to inherit either an AA or BB from the mother at SNP 2, but not AB. In the case of an unmatching copy error, the fetus inherits an AB from the mother at both SNPs. The allele distribution hypotheses made by a CNV calling method that takes into account linkage can make these predictions, and therefore correspond to the actual allele measurements to a considerably greater extent than a CNV calling method that does not take into account linkage.

In some embodiments, phased genetic data is used to determine if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of an individual (such as in the genome of one or more cells or in cfDNA or cfRNA). Exemplary overrepresentations include the duplication of the first homologous chromosome segment or the deletion of the second homologous chromosome segment. In some embodiments, there is not an overrepresentation since the first and homologous chromosome segments are present in equal proportions (such as one copy of each segment in a diploid sample). In some embodiments, calculated allele ratios in a nucleic acid sample are compared to expected allele ratios to determine if there is an overrepresentation as described further below. In this specification the phrase "a first homologous chromosome segment as compared to a second homologous chromosome segment" means a first homolog of a chromosome segment and a second homolog of the chromosome segment.

In some embodiments, the method includes obtaining phased genetic data for the first homologous chromosome segment comprising the identity of the allele present at that locus on the first homologous chromosome segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, obtaining phased genetic data for the second homologous chromosome segment comprising the identity of the allele present at that locus on the second

# US 11,530,454 B2

4

homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and obtaining measured genetic allelic data comprising, for each of the alleles at each of the loci in the set of polymorphic loci, the amount of each allele present in a sample of DNA or RNA from one or more target cells and one or more non-target cells from the individual. In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment; calcu- 10 lating, for each of the hypotheses, expected genetic data for the plurality of loci in the sample from the obtained phased genetic data for one or more possible ratios of DNA or RNA from the one or more target cells to the total DNA or RNA in the sample; calculating (such as calculating on a com- 15 puter) for each possible ratio of DNA or RNA and for each hypothesis, the data fit between the obtained genetic data of the sample and the expected genetic data for the sample for that possible ratio of DNA or RNA and for that hypothesis; ranking one or more of the hypotheses according to the data 20 fit; and selecting the hypothesis that is ranked the highest, thereby determining the degree of overrepresentation of the number of copies of the first homologous chromosome segment in the genome of one or more cells from the individual.

In one aspect, the invention features a method for determining a number of copies of a chromosome or chromosome segment of interest in the genome of a fetus. In some embodiments, the method includes obtaining phased genetic data for at least one biological parent of the fetus, wherein 30 the phased genetic data comprises the identity of the allele present for each locus in a set of polymorphic loci on a first homologous chromosome segment and a second homologous chromosome segment in the parent. In some embodiments, the method includes obtaining genetic data at the set 3: of polymorphic loci on the chromosome or chromosome segment in a mixed sample of DNA or RNA comprising fetal DNA or RNA and maternal DNA or RNA from the mother of the fetus by measuring the quantity of each allele at each locus. In some embodiments, the method includes enumer- 40 ating a set of one or more hypotheses specifying the number of copies of the chromosome or chromosome segment of interest present in the genome of the fetus. In some embodiments, the method includes creating (such as creating on a computer) for each of the hypotheses, a probability distri- 45 bution of the expected quantity of each allele at each of the plurality of loci in mixed sample from the (i) the obtained phased genetic data from the parent(s) and optionally (ii) the probability of one or more crossovers that may have occurred during the formation of a gamete that contributed 50 a copy of the chromosome or chromosome segment of interest to the fetus; calculating (such as calculating on a computer) a fit, for each of the hypotheses, between (1) the obtained genetic data of the mixed sample and (2) the probability distribution of the expected quantity of each 55 allele at each of the plurality of loci in mixed sample for that hypothesis; ranking one or more of the hypotheses according to the data fit; and selecting the hypothesis that is ranked the highest, thereby determining the number of copies of the chromosome segment of interest in the genome of the fetus. 60

In some embodiments, the method involves obtaining phased genetic data using any of the methods described herein or any known method. In some embodiments, the method involves simultaneously or sequentially in any order (i) obtaining phased genetic data for the first homologous chromosome segment comprising the identity of the allele present at that locus on the first homologous chromosome

48

segment for each locus in a set of polymorphic loci on the first homologous chromosome segment, (ii) obtaining phased genetic data for the second homologous chromosome segment comprising the identity of the allele present at that locus on the second homologous chromosome segment for each locus in the set of polymorphic loci on the second homologous chromosome segment, and (iii) obtaining measured genetic allelic data comprising the amount of each allele at each of the loci in the set of polymorphic loci in a sample of DNA from one or more cells from the individual.

In some embodiments, the method involves calculating allele ratios for one or more loci in the set of polymorphic loci that are heterozygous in at least one cell from which the sample was derived (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother). In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles divided by the total measured quantity of all the alleles for the locus. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles (such as the allele on the first homologous chromosome segment) divided by the measured quantity of one or more other alleles (such as the allele on the second homologous chromosome segment) for the locus. The calculated allele 25 ratios may be calculated using any of the methods described herein or any standard method (such as any mathematical transformation of the calculated allele ratios described herein).

In some embodiments, the method involves determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment by comparing one or more calculated allele ratios for a locus to an allele ratio that is expected for that locus if the first and second homologous chromosome segments are present in equal proportions. In some embodiments, the expected allele ratio assumes the possible alleles for a locus have an equal likelihood of being present. In some embodiments in which the calculated allele ratio for a particular locus is the measured quantity of one of the alleles divided by the total measured quantity of all the alleles for the locus, the corresponding expected allele ratio is 0.5 for a biallelic locus, or 1/3 for a triallelic locus. In some embodiments, the expected allele ratio is the same for all the loci, such as 0.5 for all loci. In some embodiments, the expected allele ratio assumes that the possible alleles for a locus can have a different likelihood of being present, such as the likelihood based on the frequency of each of the alleles in a particular population that the subject belongs in, such as a population based on the ancestry of the subject. Such allele frequencies are publicly available (see, e.g., HapMap Project; Perlegen Human Haplotype Project; web at ncbi.nlm.nih.gov/projects/SNP/; Sherry S T, Ward M H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan. 1; 29(1):308-11, which are each incorporated by reference in its entirety). In some embodiments, the expected allele ratio is the allele ratio that is expected for the particular individual being tested for a particular hypothesis specifying the degree of overrepresentation of the first homologous chromosome segment. For example, the expected allele ratio for a particular individual may be determined based on phased or unphased genetic data from the individual (such as from a sample from the individual that is unlikely to have a deletion or duplication such as a noncancerous sample) or data from one or more relatives from the individual. In some embodiments for prenatal testing, the expected allele ratio is the allele ratio that is expected for a mixed sample that includes DNA or

# US 11,530,454 B2

49

RNA from the pregnant mother and the fetus (such as a maternal plasma or serum sample that includes cfDNA from the mother and cfDNA from the fetus) for a particular hypothesis specifying the degree of overrepresentation of the first homologous chromosome segment. For example, 5 the expected allele ratio for the mixed sample may be determined based on genetic data from the mother and predicted genetic data for the fetus (such as predictions for alleles that the fetus may have inherited from the mother and/or father). In some embodiments, phased or unphased 10 genetic data from a sample of DNA or RNA from only the mother (such as the buffy coat from a maternal blood sample) is to determine the alleles from the maternal DNA or RNA in the mixed sample as well as alleles that the fetus may have been inherited from the mother (and thus may be 15 present in the fetal DNA or RNA in the mixed sample). In some embodiments, phased or unphased genetic data from a sample of DNA or RNA from only the father is used to determine the alleles that the fetus may have been inherited from the father (and thus may be present in the fetal DNA 20 or RNA in the mixed sample). The expected allele ratios may be calculated using any of the methods described herein or any standard method (such as any mathematical transformation of the expected allele ratios described herein) (U.S. Publication No 2012/0270212, filed Nov. 18, 2011, which is 25 hereby incorporated by reference in its entirety)

In some embodiments, a calculated allele ratio is indicative of an overrepresentation of the number of copies of the first homologous chromosome segment if either (i) the allele ratio for the measured quantity of the allele present at that 30 locus on the first homologous chromosome divided by the total measured quantity of all the alleles for the locus is greater than the expected allele ratio for that locus, or (ii) the allele ratio for the measured quantity of the allele present at that locus on the second homologous chromosome divided 3: by the total measured quantity of all the alleles for the locus is less than the expected allele ratio for that locus. In some embodiments, a calculated allele ratio is only considered indicative of overrepresentation if it is significantly greater or lower than the expected ratio for that locus. In some 40 embodiments, a calculated allele ratio is indicative of no overrepresentation of the number of copies of the first homologous chromosome segment if either (i) the allele ratio for the measured quantity of the allele present at that locus on the first homologous chromosome divided by the 45 total measured quantity of all the alleles for the locus is less than or equal to the expected allele ratio for that locus, or (ii) the allele ratio for the measured quantity of the allele present at that locus on the second homologous chromosome divided by the total measured quantity of all the alleles for 50 the locus is greater than or equal to the expected allele ratio for that locus. In some embodiments, calculated ratios equal to the corresponding expected ratio are ignored (since they are indicative of no overrepresentation).

In various embodiments, one or more of the following 55 methods is used to compare one or more of the calculated allele ratios to the corresponding expected allele ratio(s). In some embodiments, one determines whether the calculated allele ratio is above or below the expected allele ratio for a particular locus irrespective of the magnitude of the difference. In some embodiments, one determines the magnitude of the difference between the calculated allele ratio and the expected allele ratio for a particular locus irrespective of whether the calculated allele ratio is above or below the expected allele ratio. In some embodiments, one determines 65 whether the calculated allele ratio is above or below the expected allele ratio and the magnitude of the difference for

50

a particular locus. In some embodiments, one determines whether the average or weighted average value of the calculated allele ratios is above or below the average or weighted average value of the expected allele ratios irrespective of the magnitude of the difference. In some embodiments, one determines the magnitude of the difference between the average or weighted average value of the calculated allele ratios and the average or weighted average value of the expected allele ratios irrespective of whether the average or weighted average of the calculated allele ratio is above or below the average or weighted average value of the expected allele ratio. In some embodiments, one determines whether the average or weighted average value of the calculated allele ratios is above or below the average or weighted average value of the expected allele ratios and the magnitude of the difference. In some embodiments, one determines an average or weighted average value of the magnitude of the difference between the calculated allele ratios and the expected allele ratios.

In some embodiments, the magnitude of the difference between the calculated allele ratio and the expected allele ratio for one or more loci is used to determine whether the overrepresentation of the number of copies of the first homologous chromosome segment is due to a duplication of the first homologous chromosome segment or a deletion of the second homologous chromosome segment in the genome of one or more of the cells.

In some embodiments, an overrepresentation of the number of copies of the first homologous chromosome segment is determined to be present if one or more of following conditions is met. In some embodiments, the number of calculated allele ratios that are indicative of an overrepresentation of the number of copies of the first homologous chromosome segment is above a threshold value. In some embodiments, the number of calculated allele ratios that are indicative of no overrepresentation of the number of copies of the first homologous chromosome segment is below a threshold value. In some embodiments, the magnitude of the difference between the calculated allele ratios that are indicative of an overrepresentation of the number of copies of the first homologous chromosome segment and the corresponding expected allele ratios is above a threshold value. In some embodiments, for all calculated allele ratios that are indicative of overrepresentation, the sum of the magnitude of the difference between a calculated allele ratio and the corresponding expected allele ratio is above a threshold value. In some embodiments, the magnitude of the difference between the calculated allele ratios that are indicative of no overrepresentation of the number of copies of the first homologous chromosome segment and the corresponding expected allele ratios is below a threshold value. In some embodiments, the average or weighted average value of the calculated allele ratios for the measured quantity of the allele present on the first homologous chromosome divided by the total measured quantity of all the alleles for the locus is greater than the average or weighted average value of the expected allele ratios by at least a threshold value. In some embodiments, the average or weighted average value of the calculated allele ratios for the measured quantity of the allele present on the second homologous chromosome divided by the total measured quantity of all the alleles for the locus is less than the average or weighted average value of the expected allele ratios by at least a threshold value. In some embodiments, the data fit between the calculated allele ratios and allele ratios that are predicted for an overrepresentation of the number of copies of the first homologous chromosome segment is below a threshold value (indicative of a

# US 11,530,454 B2

51

good data fit). In some embodiments, the data fit between the calculated allele ratios and allele ratios that are predicted for no overrepresentation of the number of copies of the first homologous chromosome segment is above a threshold value (indicative of a poor data fit).

In some embodiments, an overrepresentation of the number of copies of the first homologous chromosome segment is determined to be absent if one or more of following conditions is met. In some embodiments, the number of calculated allele ratios that are indicative of an overrepre- 10 sentation of the number of copies of the first homologous chromosome segment is below a threshold value. In some embodiments, the number of calculated allele ratios that are indicative of no overrepresentation of the number of copies of the first homologous chromosome segment is above a 15 threshold value. In some embodiments, the magnitude of the difference between the calculated allele ratios that are indicative of an overrepresentation of the number of copies of the first homologous chromosome segment and the corresponding expected allele ratios is below a threshold value. 20 In some embodiments, the magnitude of the difference between the calculated allele ratios that are indicative of no overrepresentation of the number of copies of the first homologous chromosome segment and the corresponding expected allele ratios is above a threshold value. In some 25 embodiments, the average or weighted average value of the calculated allele ratios for the measured quantity of the allele present on the first homologous chromosome divided by the total measured quantity of all the alleles for the locus minus the average or weighted average value of the expected allele 30 ratios is less than a threshold value. In some embodiments, the average or weighted average value of the expected allele ratios minus the average or weighted average value of the calculated allele ratios for the measured quantity of the allele present on the second homologous chromosome divided by the total measured quantity of all the alleles for the locus is less than a threshold value. In some embodiments, the data fit between the calculated allele ratios and allele ratios that are predicted for an overrepresentation of the number of copies of the first homologous chromosome segment is 40 above a threshold value. In some embodiments, the data fit between the calculated allele ratios and allele ratios that are predicted for no overrepresentation of the number of copies of the first homologous chromosome segment is below a threshold value. In some embodiments, the threshold is 45 determined from empirical testing of samples known to have a CNV of interest and/or samples known to lack the CNV.

In some embodiments, determining if there is an overrepresentation of the number of copies of the first homologous chromosome segment includes enumerating a set of 50 one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment. On exemplary hypothesis is the absence of an overrepresentation since the first and homologous chromosome segments are present in equal proportions (such as one copy of 55 each segment in a diploid sample). Other exemplary hypotheses include the first homologous chromosome segment being duplicated one or more times (such as 1, 2, 3, 4, 5, or more extra copies of the first homologous chromosome compared to the number of copies of the second homologous 60 chromosome segment). Another exemplary hypothesis includes the deletion of the second homologous chromosome segment. Yet another exemplary hypothesis is the deletion of both the first and the second homologous chromosome segments. In some embodiments, predicted allele 65 ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or

52

heterozygous in the mother) are estimated for each hypothesis given the degree of overrepresentation specified by that hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing the calculated allele ratios to the predicted allele ratios, and the hypothesis with the greatest likelihood is selected.

In some embodiments, an expected distribution of a test statistic is calculated using the predicted allele ratios for each hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing a test statistic that is calculated using the calculated allele ratios to the expected distribution of the test statistic that is calculated using the predicted allele ratios, and the hypothesis with the greatest likelihood is selected.

In some embodiments, predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) are estimated given the phased genetic data for the first homologous chromosome segment, the phased genetic data for the second homologous chromosome segment, and the degree of overrepresentation specified by that hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing the calculated allele ratios to the predicted allele ratios; and the hypothesis with the greatest likelihood is selected.

Use of Mixed Samples

It will be understood that for many embodiments, the sample is a mixed sample with DNA or RNA from one or more target cells and one or more non-target cells. In some embodiments, the target cells are cells that have a CNV, such as a deletion or duplication of interest, and the non-target cells are cells that do not have the copy number variation of interest (such as a mixture of cells with the deletion or duplication of interest and cells without any of the deletions or duplications being tested). In some embodiments, the target cells are cells that are associated with a disease or disorder or an increased risk for disease or disorder (such as cancer cells), and the non-target cells are cells that are not associated with a disease or disorder or an increased risk for disease or disorder (such as noncancerous cells). In some embodiments, the target cells all have the same CNV. In some embodiments, two or more target cells have different CNVs. In some embodiments, one or more of the target cells has a CNV, polymorphism, or mutation associated with the disease or disorder or an increased risk for disease or disorder that is not found it at least one other target cell. In some such embodiments, the fraction of the cells that are associated with the disease or disorder or an increased risk for disease or disorder out of the total cells from a sample is assumed to be greater than or equal to the fraction of the most frequent of these CNVs, polymorphisms, or mutations in the sample. For example if 6% of the cells have a K-ras mutation, and 8% of the cells have a BRAF mutation, at least 8% of the cells are assumed to be cancerous.

In some embodiments, the ratio of DNA (or RNA) from the one or more target cells to the total DNA (or RNA) in the sample is calculated. In some embodiments, a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment are enumerated. In some embodiments, predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) are estimated given the calculated ratio of DNA or RNA and the degree of overrepresentation specified by that hypothesis are estimated for each hypothesis. In some embodiments, the likelihood that the hypothesis is correct is calculated by comparing the calculated

# US 11,530,454 B2

5

allele ratios to the predicted allele ratios, and the hypothesis with the greatest likelihood is selected.

In some embodiments, an expected distribution of a test statistic calculated using the predicted allele ratios and the calculated ratio of DNA or RNA is estimated for each 5 hypothesis. In some embodiments, the likelihood that the hypothesis is correct is determined by comparing a test statistic calculated using the calculated allele ratios and the calculated ratio of DNA or RNA to the expected distribution of the test statistic calculated using the predicted allele ratios and the calculated ratio of DNA or RNA, and the hypothesis with the greatest likelihood is selected.

In some embodiments, the method includes enumerating a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome 15 segment. In some embodiments, the method includes estimating, for each hypothesis, either (i) predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) given the degree of overrepresentation 20 specified by that hypothesis or (ii) for one or more possible ratios of DNA or RNA, an expected distribution of a test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA from the one or more target cells to the total DNA or RNA in the sample. In some 25 embodiments, a data fit is calculated by comparing either (i) the calculated allele ratios to the predicted allele ratios, or (ii) a test statistic calculated using the calculated allele ratios and the possible ratio of DNA or RNA to the expected distribution of the test statistic calculated using the predicted 30 allele ratios and the possible ratio of DNA or RNA. In some embodiments, one or more of the hypotheses are ranked according to the data fit, and the hypothesis that is ranked the highest is selected. In some embodiments, a technique or algorithm, such as a search algorithm, is used for one or 3 more of the following steps: calculating the data fit, ranking the hypotheses, or selecting the hypothesis that is ranked the highest. In some embodiments, the data fit is a fit to a beta-binomial distribution or a fit to a binomial distribution. In some embodiments, the technique or algorithm is selected 40 from the group consisting of maximum likelihood estimation, maximum a-posteriori estimation, Bayesian estimation, dynamic estimation (such as dynamic Bayesian estimation), and expectation-maximization estimation. In some embodiments, the method includes applying the technique or algo- 45 rithm to the obtained genetic data and the expected genetic

In some embodiments, the method includes creating a partition of possible ratios that range from a lower limit to an upper limit for the ratio of DNA or RNA from the one or 50 more target cells to the total DNA or RNA in the sample. In some embodiments, a set of one or more hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment are enumerated. In some embodiments, the method includes estimating, for each of 55 the possible ratios of DNA or RNA in the partition and for each hypothesis, either (i) predicted allele ratios for the loci that are heterozygous in at least one cell (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother) given the possible ratio of DNA or RNA and the 60 degree of overrepresentation specified by that hypothesis or (ii) an expected distribution of a test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA. In some embodiments, the method includes calculating, for each of the possible ratios of DNA or RNA 65 in the partition and for each hypothesis, the likelihood that the hypothesis is correct by comparing either (i) the calcu54

lated allele ratios to the predicted allele ratios, or (ii) a test statistic calculated using the calculated allele ratios and the possible ratio of DNA or RNA to the expected distribution of the test statistic calculated using the predicted allele ratios and the possible ratio of DNA or RNA. In some embodiments, the combined probability for each hypothesis is determined by combining the probabilities of that hypothesis for each of the possible ratios in the partition; and the hypothesis with the greatest combined probability is selected. In some embodiments, the combined probability for each hypothesis is determining by weighting the probability of a hypothesis for a particular possible ratio based on the likelihood that the possible ratio is the correct ratio.

In some embodiments, a technique selected from the group consisting of maximum likelihood estimation, maximum a-posteriori estimation, Bayesian estimation, dynamic estimation (such as dynamic Bayesian estimation), and expectation-maximization estimation is used to estimate the ratio of DNA or RNA from the one or more target cells to the total DNA or RNA in the sample. In some embodiments, the ratio of DNA or RNA in the sample is assumed to be the same for two or more (or all) of the CNVs of interest. In some embodiments, the ratio of DNA or RNA from the one or more target cells to the total DNA or RNA from the one or more target cells to the total DNA or RNA in the sample is calculated for each CNV of interest.

Exemplary Methods for Using Imperfectly Phased Data

It will be understood that for many embodiments, imperfectly phased data is used. For example, it may not be known with 100% certainty which allele is present for one or more of the loci on the first and/or second homologous chromosome segment. In some embodiments, the priors for possible haplotypes of the individual (such as haplotypes based on population based haplotype frequencies) are used in calculating the probability of each hypothesis. In some embodiments, the priors for possible haplotypes are adjusted by either using another method to phase the genetic data or by using phased data from other subjects (such as prior subjects) to refine population data used for informatics based phasing of the individual.

In some embodiments, the phased genetic data comprises probabilistic data for two or more possible sets of phased genetic data, wherein each possible set of phased data comprises a possible identity of the allele present at each locus in the set of polymorphic loci on the first homologous chromosome segment and a possible identity of the allele present at each locus in the set of polymorphic loci on the second homologous chromosome segment. In some embodiments, the probability for at least one of the hypotheses is determined for each of the possible sets of phased genetic data. In some embodiments, the combined probability for the hypothesis is determined by combining the probabilities of the hypothesis for each of the possible sets of phased genetic data; and the hypothesis with the greatest combined probability is selected.

Any of the methods disclosed herein or any known method may be used to generate imperfectly phased data (such as using population based haplotype frequencies to infer the most likely phase) for use in the claimed methods. In some embodiments, phased data is obtained by probabilistically combining haplotypes of smaller segments. For example, possible haplotypes can be determined based on possible combinations of one haplotype from a first region with another haplotype from another region from the same chromosome. The probability that particular haplotypes from different regions are part of the same, larger haplotype block on the same chromosome can be determined using,

5

e.g., population based haplotype frequencies and/or known recombination rates between the different regions.

In some embodiments, a single hypothesis rejection test is used for the null hypothesis of disomy. In some embodiments, the probability of the disomy hypothesis is calculated, and the hypothesis of disomy is rejected if the probability is below a given threshold value (such as less than 1 in 1,000). If the null hypothesis is rejected, this could be due to errors in the imperfectly phased data or due to the presence of a CNV. In some embodiments, more accurate 10 phased data is obtained (such as phased data from any of the molecular phasing methods disclosed herein to obtain actual phased data rather than bioinformatics-based inferred phased data). In some embodiments, the probability of the disomy hypothesis is recalculated using the more accurate 15 phased data to determine if the disomy hypothesis should still be rejected. Rejection of this hypothesis indicates that a duplication or deletion of the chromosome segment is present. If desired, the false positive rate can be altered by adjusting the threshold value.

# Further Exemplary Embodiments for Determining Ploidy Using Phased Data

In illustrative embodiments, provided herein is a method 25 for determining ploidy of a chromosomal segment in a sample of an individual. The method includes the following steps:

- a. receiving allele frequency data comprising the amount of each allele present in the sample at each loci in a set 30 of polymorphic loci on the chromosomal segment;
- b. generating phased allelic information for the set of polymorphic loci by estimating the phase of the allele frequency data:
- c. generating individual probabilities of allele frequencies 35 for the polymorphic loci for different ploidy states using the allele frequency data;
- d. generating joint probabilities for the set of polymorphic loci using the individual probabilities and the phased allelic information; and
- e. selecting, based on the joint probabilities, a best fit model indicative of chromosomal ploidy, thereby determining ploidy of the chromosomal segment.

As disclosed herein, the allele frequency data (also referred to herein as measured genetic allelic data) can be 45 generated by methods known in the art. For example, the data can be generated using qPCR or microarrays. In one illustrative embodiment, the data is generated using nucleic acid sequence data, especially high throughput nucleic acid sequence data.

In certain illustrative examples, the allele frequency data is corrected for errors before it is used to generate individual probabilities. In specific illustrative embodiments, the errors that are corrected include allele amplification efficiency bias. In other embodiments, the errors that are corrected include 55 ambient contamination and genotype contamination. In some embodiments, errors that are corrected include allele amplification bias, ambient contamination and genotype contamination.

In certain embodiments, the individual probabilities are 60 generated using a set of models of both different ploidy states and allelic imbalance fractions for the set of polymorphic loci. In these embodiments, and other embodiments, the joint probabilities are generated by considering the linkage between polymorphic loci on the chromosome segment. 65

Accordingly, in one illustrative embodiment that combines some of these embodiments, provided herein is a 56

method for detecting chromosomal ploidy in a sample of an individual, that includes the following steps:

- a. receiving nucleic acid sequence data for alleles at a set of polymorphic loci on a chromosome segment in the individual:
- b. detecting allele frequencies at the set of loci using the nucleic acid sequence data;
- c. correcting for allele amplification efficiency bias in the detected allele frequencies to generate corrected allele frequencies for the set of polymorphic loci;
- d. generating phased allelic information for the set of polymorphic loci by estimating the phase of the nucleic acid sequence data;
- e. generating individual probabilities of allele frequencies for the polymorphic loci for different ploidy states by comparing the corrected allele frequencies to a set of models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci;
- f. generating joint probabilities for the set of polymorphic loci by combining the individual probabilities considering the linkage between polymorphic loci on the chromosome segment; and
- g. selecting, based on the joint probabilities, the best fit model indicative of chromosomal aneuploidy.

As disclosed herein, the individual probabilities can be generated using a set of models or hypothesis of both different ploidy states and average allelic imbalance fractions for the set of polymorphic loci. For example, in a particularly illustrative example, individual probabilities are generated by modeling ploidy states of a first homolog of the chromosome segment and a second homolog of the chromosome segment. The ploidy states that are modeled include the following:

- (1) all cells have no deletion or amplification of the first homolog or the second homolog of the chromosome segment:
- (2) at least some cells have a deletion of the first homolog or an amplification of the second homolog of the chromosome segment; and
- (3) at least some cells have a deletion of the second homolog or an amplification of the first homolog of the chromosome segment.

It will be understood that the above models can also be referred to as hypothesis that are used to constrain a model. Therefore, demonstrated above are 3 hypothesis that can be used.

The average allelic imbalance fractions modeled can include any range of average allelic imbalance that includes the actual average allelic imbalance of the chromosomal segment. For example, in certain illustrative embodiments, the range of average allelic imbalance that is modeled can be between 0, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.75, 1, 2, 2.5, 3, 4, and 5% on the low end, and 1, 2, 2.5, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70 80 90, 95, and 99% on the high end. The intervals for the modeling with the range can be any interval depending on the computing power used and the time allowed for the analysis. For example, 0.01, 0.05, 0.02, or 0.1 intervals can be modeled.

In certain illustrative embodiments, the sample has an average allelic imbalance for the chromosomal segment of between 0.4% and 5%. In certain embodiments, the average allelic imbalance is low. In these embodiments, average allelic imbalance is typically less than 10%. In certain illustrative embodiments, the allelic imbalance is between 0.25, 0.3, 0.4, 0.5, 0.6, 0.75, 1, 2, 2.5, 3, 4, and 5% on the low end, and 1, 2, 2.5, 3, 4, and 5% on the high end. In other exemplary embodiments, the average allelic imbalance is

between 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0? on the low end and 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 3.0, 4.0, or 5.0? on the high end. For example, the average allelic imbalance of the sample in an illustrative example is between 0.45 and 2.5%. In another example, the average allelic imbalance is detected with a sensitivity of 0.45, 0.5, 0.6, 0.8, 0.8, 0.9, or 1.0. In An exemplary sample with low allelic imbalance in methods of the present invention include plasma samples from individuals with cancer having circulating tumor DNA or plasma samples from pregnant females having circulating fetal DNA.

It will be understood that for SNVs, the proportion of abnormal DNA is typically measured using mutant allele frequency (number of mutant alleles at a locus/total number of alleles at that locus). Since the difference between the amounts of two homologs in tumours is analogous, we measure the proportion of abnormal DNA for a CNV by the average allelic imbalance (AAI), defined as I(H1–H2)I/(H1+H2), where Hi is the average number of copies of homolog to homolog ratio, of homolog i. The maximum homolog ratio is the homolog ratio of the more abundant homolog.

Assay drop-out rate is the percentage of SNPs with no reads, estimated using all SNPs. Single allele drop-out (ADO) rate is the percentage of SNPs with only one allele present, estimated using only heterozygous SNPs. Genotype confidence can be determined by fitting a binomial distribution to the number of reads at each SNP that were B-allele reads, and using the ploidy status of the focal region of the SNP to estimate the probability of each genotype.

For tumor tissue samples, chromosomal aneuploidy (exemplified in this paragraph by CNVs) can be delineated by transitions between allele frequency distributions. In plasma samples, CNVs can be identified by a maximum likelihood algorithm that searches for plasma CNVs in regions where the tumor sample from the same individual also has CNVs, using haplotype information deduced from the tumor sample. This algorithm can model expected allelic frequencies across all allelic imbalance ratios at 0.025% intervals for three sets of hypotheses: (1) all cells are normal (no allelic imbalance), (2) some/all cells have a homolog 1 deletion or homolog 2 amplification, or (3) some/all cells have a homolog 2 deletion or homolog 1 amplification. The likelihood of each hypothesis can be determined at each SNP using a Bayesian classifier based on a beta binomial model of expected and observed allele frequencies at all heterozygous SNPs, and then the joint likelihood across multiple SNPs can be calculated, in certain illustrative embodiments taking linkage of the SNP loci into consider-50 ation, as exemplified herein. The maximum likelihood hypothesis can then be selected.

Consider a chromosomal region with an average of N copies in the tumor, and let c denote the fraction of DNA in plasma derived from the mixture of normal and tumour cells in a disomic region. AAI is calculated as:

$$AAI = \frac{c|N-2|}{2+c(N-2)} -$$

In certain illustrative examples, the allele frequency data is corrected for errors before it is used to generate individual probabilities. Different types of error and/or bias correction are disclosed herein. In specific illustrative embodiments, 65 the errors that are corrected are allele amplification efficiency bias. In other embodiments, the errors that are

58

corrected include ambient contamination and genotype contamination. In some embodiments, errors that are corrected include allele amplification bias, ambient contamination and genotype contamination.

It will be understood that allele amplification efficiency bias can be determined for an allele as part of an experiment or laboratory determination that includes an on test sample, or it can be determined at a different time using a set of samples that include the allele whose efficiency is being calculated. Ambient contamination and genotype contamination are typically determined on the same run as the on-test sample analysis.

In certain embodiments, ambient contamination and genotype contamination are determined for homozygous alleles in the sample. It will be understood that for any given sample from an individual some loci in the sample, will be heterozygous and others will be homozygous, even if a locus is selected for analysis because it has a relatively high heterozygosity in the population. It is advantageous in some embodiments, although ploidy of a chromosomal segment may be determined using heterozygous loci for an individual, homozygous loci can be used to calculate ambient and genotype contamination.

In certain illustrative examples, the selecting is performed by analyzing a magnitude of a difference between the phased allelic information and estimated allelic frequencies generated for the models.

In illustrative examples, the individual probabilities of allele frequencies are generated based on a beta binomial model of expected and observed allele frequencies at the set of polymorphic loci. In illustrative examples, the individual probabilities are generated using a Bayesian classifier.

In certain illustrative embodiments, the nucleic acid sequence data is generated by performing high throughput DNA sequencing of a plurality of copies of a series of amplicons generated using a multiplex amplification reaction, wherein each amplicon of the series of amplicons spans at least one polymorphic loci of the set of polymorphic loci and wherein each of the polymeric loci of the set is amplification reaction is performed under limiting primer conditions for at least ½ of the reactions. In some embodiments, limiting primer concentrations are used in ½10, ½5, ¼4, ½3, ½2, or all of the reactions of the multiplex reaction. Provided herein are factors to consider to achieve limiting primer conditions in an amplification reaction such as PCR.

In certain embodiments, methods provided herein detect ploidy for multiple chromosomal segments across multiple chromosomes. Accordingly, the chromosomal ploidy in these embodiments is determined for a set of chromosome segments in the sample. For these embodiments, higher multiplex amplification reactions are needed. Accordingly, for these embodiments the multiplex amplification reaction can include, for example, between 2,500 and 50,000 multiplex reactions. In certain embodiments, the following ranges of multiplex reactions are performed: between 100, 200, 250, 500, 1000, 2500, 5000, 10,000, 20,000, 25000, 50000 on the low end of the range and between 200, 250, 500, 1000, 2500, 5000, 10,000, 20,000, 25000, 50000, and 100, 000 on the high end of the range.

In illustrative embodiments, the set of polymorphic loci is a set of loci that are known to exhibit high heterozygosity. However, it is expected that for any given individual, some of those loci will be homozygous. In certain illustrative embodiments, methods of the invention utilize nucleic acid sequence information for both homozygous and heterozygous loci for an individual. The homozygous loci of an

individual are used, for example, for error correction, whereas heterozygous loci are used for the determination of allelic imbalance of the sample. In certain embodiments, at least 10% of the polymorphic loci are heterozygous loci for the individual.

As disclosed herein, preference is given for analyzing target SNP loci that are known to be heterozygous in the population. Accordingly, in certain embodiments, polymorphic loci are chosen wherein at least 10, 20, 25, 50, 75, 80, 90, 95, 99, or 100% of the polymorphic loci are known to be 10 heterozygous in the population.

As disclosed herein, in certain embodiments the sample is a plasma sample from a pregnant female.

In some examples, the method further comprises performing the method on a control sample with a known average allelic imbalance ratio. The control can have an average allelic imbalance ratio for a particular allelic state indicative of aneuploidy of the chromosome segment, of between 0.4 and 10% to mimic an average allelic imbalance of an allele in a sample that is present in low concentrations, such as 20 would be expected for a circulating free DNA from a fetus or from a tumor.

In some embodiments, PlasmArt controls, as disclosed herein, are used as the controls. Accordingly, in certain aspects the is a sample generated by a method comprising 25 fragmenting a nucleic acid sample known to exhibit a chromosomal aneuploidy into fragments that mimic the size of fragments of DNA circulating in plasma of the individual. In certain aspects a control is used that has no aneuploidy for the chromosome segment.

In illustrative embodiments, data from one or more controls can be analyzed in the method along with a test sample. The controls for example, can include a different sample from the individual that is not suspected of containing Chromosomal aneuploidy, or a sample that is suspected of 35 containing CNV or a chromosomal aneuploidy. For example, where a test sample is a plasma sample suspected of containing circulating free tumor DNA, the method can be also be performed for a control sample from a tumor from the subject along with the plasma sample. As disclosed 40 herein, the control sample can be prepared by fragmenting a DNA sample known to exhibit a chromosomal aneuploidy. Such fragmenting can result in a DNA sample that mimics the DNA composition of an apoptotic cell, especially when the sample is from an individual afflicted with cancer. Data 45 from the control sample will increase the confidence of the detection of Chromosomal aneuploidy

In certain embodiments of the methods of determining ploidy, the sample is a plasma sample from an individual suspected of having cancer. In these embodiments, the 50 method further comprises determining based on the selecting whether copy number variation is present in cells of a tumor of the individual. For these embodiments, the sample can be a plasma sample from an individual. For these embodiments, the method can further include determining, 55 based on the selecting, whether cancer is present in the individual.

These embodiments for determining ploidy of a chromosomal segment, can further include detecting a single nucleotide variant at a single nucleotide variance location in 60 a set of single nucleotide variance locations, wherein detecting either a chromosomal aneuploidy or the single nucleotide variant or both, indicates the presence of circulating tumor nucleic acids in the sample.

These embodiments can further include receiving haplotype information of the chromosome segment for a tumor of the individual and using the haplotype information to gen60

erate the set of models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci.

As disclosed herein, certain embodiments of the methods of determining ploidy can further include removing outliers from the initial or corrected allele frequency data before comparing the initial or the corrected allele frequencies to the set of models. For example, in certain embodiments, loci allele frequencies that are at least 2 or 3 standard deviations above or below the mean value for other loci on the chromosome segment, are removed from the data before being used for the modeling.

As mentioned herein, it will be understood that for many of the embodiments provided herein, including those for determining ploidy of a chromosomal segment, imperfectly or perfectly phased data is preferably used. It will also be understood, that provided herein are a number of features that provide improvements over prior methods for detecting ploidy, and that many different combinations of these features could be used.

In certain embodiments, as illustrated in FIGS. **69-70**, provided herein are computer systems and computer readable media to perform any methods of the present invention. These include systems and computer readable media for performing methods of determining ploidy. Accordingly, and as non-limiting examples of system embodiments, to demonstrate that any of the methods provided herein can be performed using a system and a computer readable medium using the disclosure herein, in another aspect, provided herein is a system for detecting chromosomal ploidy in a sample of an individual, the system comprising:

- a. an input processor configured to receive allelic frequency data comprising the amount of each allele present in the sample at each loci in a set of polymorphic loci on the chromosomal segment;
- b. a modeler configured to:
  - generate phased allelic information for the set of polymorphic loci by estimating the phase of the allele frequency data; and
  - ii. generate individual probabilities of allele frequencies for the polymorphic loci for different ploidy states using the allele frequency data; and
  - iii. generate joint probabilities for the set of polymorphic loci using the individual probabilities and the phased allelic information; and
- c. a hypothesis manager configured to select, based on the joint probabilities, a best fit model indicative of chromosomal ploidy, thereby determining ploidy of the chromosomal segment.

In certain embodiments of this system embodiment, the allele frequency data is data generated by a nucleic acid sequencing system. In certain embodiments, the system further comprises an error correction unit configured to correct for errors in the allele frequency data, wherein the corrected allele frequency data is used by the modeler for to generate individual probabilities. In certain embodiments the error correction unit corrects for allele amplification efficiency bias. In certain embodiments, the modeler generates the individual probabilities using a set of models of both different ploidy states and allelic imbalance fractions for the set of polymorphic loci. The modeler, in certain exemplary embodiments generates the joint probabilities by considering the linkage between polymorphic loci on the chromosome segment.

In one illustrative embodiment, provided herein is a system for detecting chromosomal ploidy in a sample of an individual, that includes the following:

- a. an input processor configured to receive nucleic acid sequence data for alleles at a set of polymorphic loci on a chromosome segment in the individual and detect allele frequencies at the set of loci using the nucleic acid sequence data:
- b. an error correction unit configured to correct for errors in the detected allele frequencies and generate corrected allele frequencies for the set of polymorphic
- c. a modeler configured to:
  - i. generate phased allelic information for the set of polymorphic loci by estimating the phase of the nucleic acid sequence data;
  - ii. generate individual probabilities of allele frequencies for the polymorphic loci for different ploidy 15 states by comparing the phased allelic information to a set of models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci;
  - iii. generate joint probabilities for the set of polymor- 20 phic loci by combining the individual probabilities considering the relative distance between polymorphic loci on the chromosome segment; and
- d. a hypothesis manager configured to select, based on the joint probabilities, a best fit model indicative of chro- 25 mosomal aneuploidy.

In certain exemplary system embodiments provided herein the set of polymorphic loci comprises between 1000 and 50,000 polymorphic loci. In certain exemplary system embodiments provided herein the set of polymorphic loci 30 comprises 100 known heterozygosity hot spot loci. In certain exemplary system embodiments provided herein the set of polymorphic loci comprise 100 loci that are at or within 0.5 kb of a recombination hot spot.

In certain exemplary system embodiments provided 35 herein the best fit model analyzes the following ploidy states of a first homolog of the chromosome segment and a second homolog of the chromosome segment:

- (1) all cells have no deletion or amplification of the first homolog or the second homolog of the chromosome seg- 40
- (2) some or all cells have a deletion of the first homolog or an amplification of the second homolog of the chromosome segment; and
- (3) some or all cells have a deletion of the second 45 homolog or an amplification of the first homolog of the chromosome segment

In certain exemplary system embodiments provided herein the errors that are corrected comprise allelic amplification efficiency bias, contamination, and/or sequencing 50 errors. In certain exemplary system embodiments provided herein the contamination comprises ambient contamination and genotype contamination. In certain exemplary system embodiments provided herein the ambient contamination and genotype contamination is determined for homozygous 55 alleles.

In certain exemplary system embodiments provided herein the hypothesis manager is configured to analyze a magnitude of a difference between the phased allelic information and estimated allelic frequencies generated for the 60 models. In certain exemplary system embodiments provided herein the modeler generates individual probabilities of allele frequencies based on a beta binomial model of expected and observed allele frequencies at the set of polymorphic loci. In certain exemplary system embodiments 65 allele frequency data is generated from nucleic acid provided herein the modeler generates individual probabilities using a Bayesian classifier.

In certain exemplary system embodiments provided herein the nucleic acid sequence data is generated by performing high throughput DNA sequencing of a plurality of copies of a series of amplicons generated using a multiplex amplification reaction, wherein each amplicon of the series of amplicons spans at least one polymorphic loci of the set of polymorphic loci and wherein each of the polymeric loci of the set is amplified. In certain exemplary system embodiments provided herein, wherein the multiplex amplification reaction is performed under limiting primer conditions for at least ½ of the reactions. In certain exemplary system embodiments provided herein, wherein the sample has an average allelic imbalance of between 0.4% and 5%.

In certain exemplary system embodiments provided herein, the sample is a plasma sample from an individual suspected of having cancer, and the hypothesis manager is further configured to determine, based on the best fit model, whether copy number variation is present in cells of a tumor of the individual.

In certain exemplary system embodiments provided herein the sample is a plasma sample from an individual and the hypothesis manager is further configured to determine, based on the best fit model, that cancer is present in the individual. In these embodiments, the hypothesis manager can be further configured to detect a single nucleotide variant at a single nucleotide variance location in a set of single nucleotide variance locations, wherein detecting either a chromosomal aneuploidy or the single nucleotide variant or both, indicates the presence of circulating tumor nucleic acids in the sample.

In certain exemplary system embodiments provided herein, the input processor is further configured to receiving haplotype information of the chromosome segment for a tumor of the individual, and the modeler is configured to use the haplotype information to generate the set of models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci.

In certain exemplary system embodiments provided herein, the modeler generates the models over allelic imbalance fractions ranging from 0% to 25%

It will be understood that any of the methods provided herein can be executed by computer readable code that is stored on nontransitory computer readable medium. Accordingly, provided herein in one embodiment, is a nontransitory computer readable medium for detecting chromosomal ploidy in a sample of an individual, comprising computer readable code that, when executed by a processing device, causes the processing device to:

- a. receive allele frequency data comprising the amount of each allele present in the sample at each loci in a set of polymorphic loci on the chromosomal segment;
- b. generate phased allelic information for the set of polymorphic loci by estimating the phase of the allele frequency data;
- c. generate individual probabilities of allele frequencies for the polymorphic loci for different ploidy states using the allele frequency data;
- d. generate joint probabilities for the set of polymorphic loci using the individual probabilities and the phased allelic information; and
- e. select, based on the joint probabilities, a best fit model indicative of chromosomal ploidy, thereby determining ploidy of the chromosomal segment.

In certain computer readable medium embodiments, the sequence data. certain computer readable medium embodiments further comprise correcting for errors in the allele

6

frequency data and using the corrected allele frequency data for the generating individual probabilities step. In certain computer readable medium embodiments the errors that are corrected are allele amplification efficiency bias. In certain computer readable medium embodiments the individual probabilities are generated using a set of models of both different ploidy states and allelic imbalance fractions for the set of polymorphic loci. In certain computer readable medium embodiments the joint probabilities are generated by considering the linkage between polymorphic loci on the 10 chromosome segment.

In one particular embodiment, provided herein is a non-transitory computer readable medium for detecting chromosomal ploidy in a sample of an individual, comprising computer readable code that, when executed by a processing device, causes the processing device to:

- a. receive nucleic acid sequence data for alleles at a set of polymorphic loci on a chromosome segment in the individual;
- b. detect allele frequencies at the set of loci using the <sup>20</sup> nucleic acid sequence data;
- c. correcting for allele amplification efficiency bias in the detected allele frequencies to generate corrected allele frequencies for the set of polymorphic loci;
- d. generate phased allelic information for the set of <sup>25</sup> polymorphic loci by estimating the phase of the nucleic acid sequence data;
- e. generate individual probabilities of allele frequencies for the polymorphic loci for different ploidy states by comparing the corrected allele frequencies to a set of 30 models of different ploidy states and allelic imbalance fractions of the set of polymorphic loci;
- f. generate joint probabilities for the set of polymorphic loci by combining the individual probabilities considering the linkage between polymorphic loci on the 35 chromosome segment; and
- g. select, based on the joint probabilities, the best fit model indicative of chromosomal aneuploidy.

In certain illustrative computer readable medium embodiments, the selecting is performed by analyzing a magnitude of a difference between the phased allelic information and estimated allelic frequencies generated for the models.

In certain illustrative computer readable medium embodiments the individual probabilities of allele frequencies are generated based on a beta binomial model of expected and 45 observed allele frequencies at the set of polymorphic loci.

It will be understood that any of the method embodiments provided herein can be performed by executing code stored on nontransitory computer readable medium.

# Exemplary Embodiments for Detecting Cancer

In certain aspects, the present invention provides a method for detecting cancer. The sample, it will be understood can be a tumor sample or a liquid sample, such as 55 plasma, from an individual suspected of having cancer. The methods are especially effective at detecting genetic mutations such as single nucleotide alterations such as SNVs, or copy number alterations, such as CNVs in samples with low levels of these genetic alterations as a fraction of the total 60 DNA in a sample. Thus the sensitivity for detecting DNA or RNA from a cancer in samples is exceptional. The methods can combine any or all of the improvements provided herein for detecting CNV and SNV to achieve this exceptional sensitivity.

Accordingly, in certain embodiments provided herein, is a method for determining whether circulating tumor nucleic 64

acids are present in a sample in an individual, and a nontransitory computer readable medium comprising computer readable code that, when executed by a processing device, causes the processing device to carry out the method. The method includes the following steps:

- c. analyzing the sample to determine a ploidy at a set of polymorphic loci on a chromosome segment in the individual; and
- d. determining the level of average allelic imbalance present at the polymorphic loci based on the ploidy determination, wherein an average allelic imbalance equal to or greater than 0.4%, 0.45%, 0.5%, 0.6%, 0.7%, 0.75%, 0.8%, 0.9%, or 1% is indicative of the presence of circulating tumor nucleic acids, such as ctDNA, in the sample.

In certain illustrative examples, an average allelic imbalance greater than 0.4, 0.45, or 0.5% is indicative the presence of ctDNA. In certain embodiments the method for determining whether circulating tumor nucleic acids are present, further comprises detecting a single nucleotide variant at a single nucleotide variance site in a set of single nucleotide variance locations, wherein detecting either an allelic imbalance equal to or greater than 0.5% or detecting the single nucleotide variant, or both, is indicative of the presence of circulating tumor nucleic acids in the sample. It will be understood that any of the methods provided for detecting chromosomal ploidy or CNV can be used to determine the level of allelic imbalance, typically expressed as average allelic imbalance. It will be understood that any of the methods provided herein for detecting an SNV can be used to detect the single nucleotide for this aspect of the present invention.

In certain embodiments the method for determining whether circulating tumor nucleic acids are present, further comprises performing the method on a control sample with a known average allelic imbalance ratio. The control, for example, can be a sample from the tumor of the individual. In some embodiments, the control has an average allelic imbalance expected for the sample under analysis. For example, an AAI between 0.5% and 5% or an average allelic imbalance ratio of 0.5%.

In certain embodiments analyzing step in the method for determining whether circulating tumor nucleic acids are present, includes analyzing a set of chromosome segments known to exhibit aneuploidy in cancer. In certain embodiments analyzing step in the method for determining whether circulating tumor nucleic acids are present, includes analyzing between 1,000 and 50,000 or between 100 and 1000, polymorphic loci for ploidy. In certain embodiments analyzing step in the method for determining whether circulating tumor nucleic acids are present, includes analyzing between 100 and 1000 single nucleotide variant sites. For example, in these embodiments the analyzing step can include performing a multiplex PCR to amplify amplicons across the 1000 to 50,000 polymeric loci and the 100 to 1000 single nucleotide variant sites. This multiplex reaction can be set up as a single reaction or as pools of different subset multiplex reactions. The multiplex reaction methods provided herein, such as the massive multiplex PCR disclosed herein provide an exemplary process for carrying out the amplification reaction to help attain improved multiplexing and therefore, sensitivity levels.

In certain embodiments, the multiplex PCR reaction is carried out under limiting primer conditions for at least 10%, 20%, 25%, 50%, 75%, 90%, 95%, 98%, 99%, or 100% of the reactions. Improved conditions for performing the massive multiplex reaction provided herein can be used.

6

In certain aspects, the above method for determining whether circulating tumor nucleic acids are present in a sample in an individual, and all embodiments thereof, can be carried out with a system. The disclosure provides teachings regarding specific functional and structural features to carry out the methods. As a non-limiting example, the system includes the following:

- a. An input processor configured to analyze data from the sample to determine a ploidy at a set of polymorphic loci on a chromosome segment in the individual; and
- b. A modeler configured to determine the level of allelic imbalance present at the polymorphic loci based on the ploidy determination, wherein an allelic imbalance equal to or greater than 0.5% is indicative of the presence of circulating.

### Exemplary Embodiments for Detecting Single Nucleotide Variants

In certain aspects, provided herein are methods for detecting single nucleotide variants in a sample. The improved methods provided herein can achieve limits of detection of 0.015, 0.017, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4 or 0.5 percent SNV in a sample. All the embodiments for detecting SNVs can be 25 carried out with a system. The disclosure provides teachings regarding specific functional and structural features to carry out the methods. Furthermore, provided herein are embodiments comprising a nontransitory computer readable medium comprising computer readable code that, when 30 executed by a processing device, causes the processing device to carry out the methods for detecting SNVs provided herein.

Accordingly, provided herein in one embodiment, is a method for determining whether a single nucleotide variant 3 is present at a set of genomic positions in a sample from an individual, the method comprising:

- a. for each genomic position, generating an estimate of efficiency and a per cycle error rate for an amplicon spanning that genomic position, using a training data 40 set:
- b. receiving observed nucleotide identity information for each genomic position in the sample;
- c. determining a set of probabilities of single nucleotide variant percentage resulting from one or more real 45 mutations at each genomic position, by comparing the observed nucleotide identity information at each genomic position to a model of different variant percentages using the estimated amplification efficiency and the per cycle error rate for each genomic position 50 independently; and
- d. determining the most-likely real variant percentage and confidence from the set of probabilities for each genomic position.

In illustrative embodiments of the method for determining 55 whether a single nucleotide variant is present, the estimate of efficiency and the per cycle error rate is generated for a set of amplicons that span the genomic position. For example, 2, 3, 4, 5, 10, 15, 20, 25, 50, 100 or more amplicons can be included that span the genomic position.

In illustrative embodiments of the method for determining whether a single nucleotide variant is present, the observed nucleotide identity information comprises an observed number of total reads for each genomic position and an observed number of variant allele reads for each genomic position.

In illustrative embodiments of the method for determining whether a single nucleotide variant is present, the sample is 66

a plasma sample and the single nucleotide variant is present in circulating tumor DNA of the sample.

In another embodiment provided herein is a method for estimating the percent of single nucleotide variants that are present in a sample from an individual. The method includes the following steps:

- a. at a set of genomic positions, generating an estimate of efficiency and a per cycle error rate for one or more amplicon spanning those genomic positions, using a training data set;
- receiving observed nucleotide identity information for each genomic position in the sample;
- c. generating an estimated mean and variance for the total number of molecules, background error molecules and real mutation molecules for a search space comprising an initial percentage of real mutation molecules using the amplification efficiency and the per cycle error rate of the amplicons; and
- d. determining the percentage of single nucleotide variants present in the sample resulting from real mutations by determining a most-likely real single nucleotide variant percentage by fitting a distribution using the estimated means and variances to an observed nucleotide identity information in the sample.

In illustrative examples of this method for estimating the percent of single nucleotide variants that are present in a sample, the sample is a plasma sample and the single nucleotide variant is present in circulating tumor DNA of the sample.

The training data set for this embodiment of the invention typically includes samples from one or preferably a group of healthy individuals. In certain illustrative embodiments, the training data set is analyzed on the same day or even on the same run as one or more on-test samples. For example, samples from a group of 2, 3, 4, 5, 10, 15, 20, 25, 30, 36, 48, 96, 100, 192, 200, 250, 500, 1000 or more healthy individuals can be used to generate the training data set. Where data is available for larger number of healthy individuals, e.g. 96 or more, confidence increases for amplification efficiency estimates even if runs are performed in advance of performing the method for on-test samples. The PCR error rate can use nucleic acid sequence information generated not only for the SNV base location, but for the entire amplified region around the SNV, since the error rate is per amplicon. For example, using samples from 50 individuals and sequencing a 20 base pair amplicon around the SNV, error frequency data from 1000 base reads can be used to determine error frequency rate.

Typically the amplification efficiency is estimating by estimating a mean and standard deviation for amplification efficiency for an amplified segment and then fitting that to a distribution model, such as a binomial distribution or a beta binomial distribution. Error rates are determined for a PCR reaction with a known number of cycles and then a per cycle error rate is estimated.

In certain illustrative embodiments, estimating the starting molecules of the test data set further includes updating the estimate of the efficiency for the testing data set using the starting number of molecules estimated in step (b) if the observed number of reads is significantly different than the estimated number of reads. Then the estimate can be updated for a new efficiency and/or starting molecules.

The search space used for estimating the total number of molecules, background error molecules and real mutation molecules can include a search space from 0.1%, 0.2%, 0.25%, 0.5%, 1%, 2.5%, 5%, 10%, 15%, 20%, or 25% on the low end and 1%, 2%, 2.5%, 5%, 10%, 12.5%, 15%, 20%,

6

25%, 50%, 75%, 90%, or 95% on the high end copies of a base at an SNV position being the SNV base. Lower ranges, 0.1%, 0.2%, 0.25%, 0.5%, or 1% on the low end and 1%, 2%, 2.5%, 5%, 10%, 12.5%, or 15% on the high end can be used in illustrative examples for plasma samples where the method is detecting circulating tumor DNA. Higher ranges are used for tumor samples.

A distribution is fit to the number of total error molecules (background error and real mutation) in the total molecules to calculate the likelihood or probability for each possible real mutation in the search space. This distribution could be a binomial distribution or a beta binomial distribution.

The most likely real mutation is determined by determining the most likely real mutation percentage and calculating  $_{15}$ the confidence using the data from fitting the distribution. As an illustrative example and not intended to limit the clinical interpretation of the methods provided herein, if the mean mutation rate is high then the percent confidence needed to make a positive determination of an SNV is lower. For 20 example, if the mean mutation rate for an SNV in a sample using the most likely hypothesis is 5% and the percent confidence is 99%, then a positive SNV call would be made. On the other hand for this illustrative example, if the mean mutation rate for an SNV in a sample using the most likely 25 hypothesis is 1% and the percent confidence is 50%, then in certain situations a positive SNV call would not be made. It will be understood that clinical interpretation of the data would be a function of sensitivity, specificity, prevalence rate, and alternative product availability.

In one illustrative embodiment, the sample is a circulating DNA sample, such as a circulating tumor DNA sample.

In another embodiment, provided herein is a method for detecting one or more single nucleotide variants in a test sample from an individual. The method according to this 35 embodiment, includes the following steps:

- d. determining a median variant allele frequency for a plurality of control samples from each of a plurality of normal individuals, for each single nucleotide variant position in a set of single nucleotide variance positions based on results generated in a sequencing run, to identify selected single nucleotide variant positions having variant median allele frequencies in normal samples below a threshold value and to determine background error for each of the single nucleotide 45 variant positions after removing outlier samples for each of the single nucleotide variant positions;
- e. determining an observed depth of read weighted mean and variance for the selected single nucleotide variant positions for the test sample based on data generated in 50 the sequencing run for the test sample; and
- f. identifying using a computer, one or more single nucleotide variant positions with a statistically significant depth of read weighted mean compared to the background error for that position, thereby detecting 55 the one or more single nucleotide variants.

In certain embodiments of this method for detecting one or more SNVs the sample is a plasma sample, the control samples are plasma samples, and the detected one or more single nucleotide variants detected is present in circulating 60 tumor DNA of the sample. In certain embodiments of this method for detecting one or more SNVs the plurality of control samples comprises at least 25 samples. In certain illustrative embodiments, the plurality of control samples is at least 5, 10, 15, 20, 25, 50, 75, 100, 200, or 250 samples 65 on the low end and 10, 15, 20, 25, 50, 75, 100, 200, 250, 500, and 1000 samples on the high end.

68

In certain embodiments of this method for detecting one or more SNVs, outliers are removed from the data generated in the high throughput sequencing run to calculate the observed depth of read weighted mean and observed variance are determined. In certain embodiments of this method for detecting one or more SNVs the depth of read for each single nucleotide variant position for the test sample is at least 100 reads.

In certain embodiments of this method for detecting one or more SNVs the sequencing run comprises a multiplex amplification reaction performed under limited primer reaction conditions. Improved methods for performing multiplex amplification reactions provided herein, are used to perform these embodiments in illustrative examples.

Not to be limited by theory, methods of the present embodiment utilize a background error model using normal plasma samples, that are sequenced on the same sequencing run as an on-test sample, to account for run-specific artifacts. Noisy positions with normal median variant allele frequencies above a threshold, for example >0.1%, 0.2%, 0.25%, 0.5% 0.75%, and 1.0%, are removed.

Outlier samples are iteratively removed from the model to account for noise and contamination. For each base substitution of every genomic loci, the depth of read weighted mean and standard deviation of the error are calculated. In certain illustrative embodiments, samples, such as tumor or cell-free plasma samples, with single nucleotide variant positions with at least a threshold number of reads, for example, at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 250, 500, or 1000 variant reads and al Z-score greater than 2.5, 5, 7.5 or 10 against the background error model in certain embodiments, are counted as a candidate mutation.

In certain embodiments, a depth of read of greater than 100, 250, 500, 1,000, 2000, 2500, 5000, 10,000, 20,000, 25,000, 50,000, or 100,000 on the low end of the range and 2000, 2500, 5,000, 7,500, 10,000, 25,000, 50,000, 100,000, 250,000 or 500,000 reads on the high end, is attained in the sequencing run for each single nucleotide variant position in the set of single nucleotide variant positions. Typically, the sequencing run is a high throughput sequencing run. The mean or median values generated for the on-test samples, in illustrative embodiments are weighted by depth of reads. Therefore, the likelihood that a variant allele determination is real in a sample with 1 variant allele detected in 1000 reads is weighed higher than a sample with 1 variant allele detected in 10,000 reads. Since determinations of a variant allele (i.e. mutation) are not made with 100% confidence, the identified single nucleotide variant can be considered a candidate variant or a candidate mutations.

Exemplary Test Statistic for Analysis of Phased Data

An exemplary test statistic is described below for analysis of phased data from a sample known or suspected of being a mixed sample containing DNA or RNA that originated from two or more cells that are not genetically identical. Let f denote the fraction of DNA or RNA of interest, for example the fraction of DNA or RNA with a CNV of interest, or the fraction of DNA or RNA from cells of interest, such as cancer cells. In some embodiments for prenatal testing, f denotes the fraction of fetal DNA, RNA, or cells in a mixture of fetal and maternal DNA, RNA, or cells. Note that this refers to the fraction of DNA from cells of interest assuming two copies of DNA are given by each cell of interest. This differs from the DNA fraction from cells of interest at a segment that is deleted or duplicated.

The possible allelic values of each SNP are denoted A and B. AA, AB, BA, and BB are used to denote all possible ordered allele pairs. In some embodiments, SNPs with

69

ordered alleles AB or BA are analyzed. Let  $N_i$  denote the number of sequence reads of the ith SNP, and  $A_i$  and  $B_i$  denote the number of reads of the ith SNP that indicate allele A and B, respectively. It is assumed:

$$N_i=A_i+B_i$$

The allele ratio  $R_i$  is defined:

$$R_i \stackrel{\Delta}{=} \frac{A_i}{N_i}$$

Let T denote the number of SNPs targeted.

Without loss of generality, some embodiments focus on a single chromosome segment. As a matter of further clarity, in this specification the phrase "a first homologous chromosome segment as compared to a second homologous chromosome segment" means a first homolog of a chromosome segment and a second homolog of the chromosome segment. In some such embodiments, all of the target SNPs are contained in the segment chromosome of interest. In other embodiments, multiple chromosome segments are analyzed for possible copy number variations.

### MAP Estimation

This method leverages the knowledge of phasing via ordered alleles to detect the deletion or duplication of the target segment. For each SNP i, define

$$X_i \stackrel{\Delta}{=} \begin{cases} 1 & R_i < 0.5 \text{ and } SNP \ i \ AB \\ 0 & R_i \geq 0.5 \text{ and } SNP \ i \ AB \\ 0 & R_i < 0.5 \text{ and } SNP \ i \ BA \\ 1 & R_i \geq 0.5 \text{ and } SNP \ i \ BA \end{cases}$$

Then define

$$S \stackrel{\Delta}{=} \sum_{A \in SND_0} X_i$$
.

The distributions of the  $X_i$  and S under various copy number hypotheses (such as hypotheses for disomy, deletion of the first or second homolog, or duplication of the first or second homolog) are described below.

### Disomy Hypothesis

Under the hypothesis that the target segment is not deleted or duplicated,

$$X_{i} = \begin{cases} 0 & wp \ 1 - p\left(\frac{1}{2}, N_{i}\right) \\ 1 & wp \ p\left(\frac{1}{2}, N_{i}\right) \end{cases} \text{ where}$$

$$p(b, n) \stackrel{\triangle}{=} Pr\left\{X \sim Bino(b, n) \ge \frac{n}{2}\right\}.$$

If we assume a constant depth of read N, this gives us a  $_{60}$  Binomial distribution S with parameters

$$p(\frac{1}{2},N)$$
 and  $T$ .

### Deletion Hypotheses

Under the hypothesis that the first homolog is deleted (i.e., 65 an AB SNP becomes B, and a BA SNP becomes A), then  $R_i$  has a Binomial distribution with parameters

70

$$1-\frac{1}{2-f}$$

5 and T for AB SNPs, and

$$\frac{1}{2-f}$$

and T for BA SNPs. Therefore,

$$X_{i} = \begin{cases} 0 & wp \ 1 - p\left(\frac{1}{2 - f}, N_{i}\right) \\ 1 & wp \ p\left(\frac{1}{2 - f}, N_{i}\right) \end{cases}$$

20 If we assume a constant depth of read N, this gives a Binomial distribution S with parameters

$$p\left(\frac{1}{2-f}, N\right)$$
 and  $T$ .

Under the hypothesis that the second homolog is deleted (i.e., an AB SNP becomes A, and a BA SNP becomes B), then  $R_i$  has a Binomial distribution with parameter

$$\frac{1}{2-f}$$

and T for AM SNPs, and

$$1 - \frac{1}{2 - f}$$

and T for BA SNPs. Therefore,

$$X_{i} = \begin{cases} 0 & wp \ p\left(\frac{1}{2-f}, N_{i}\right) \\ 1 & wp \ 1 - p\left(\frac{1}{2-f}, N_{i}\right) \end{cases}$$

If we assume a constant depth of read N, this gives a Binomial distribution S with parameters

$$1 - p\left(\frac{1}{2-f}, N\right)$$
 and  $T$ .

### Duplication Hypotheses

Under the hypothesis that the first homolog is duplicated (i.e., an AB SNP becomes AAB, and a BA SNP becomes BBA), then  $R_i$  has a Binomial distribution with parameters

$$\frac{1+f}{2+f}$$

71

and T for AB SNPs, and

$$1 - \frac{1+f}{2+f}$$

and T for BA SNPs. Therefore,

$$X_{i} = \begin{cases} 0 & wpp\left(\frac{1+f}{2+f}, N_{i}\right) \\ 1 & wp1 - p\left(\frac{1+f}{2+f}, N_{i}\right) \end{cases}$$

If we assume a constant depth of read N, this gives us a Binomial distribution S with parameters

$$1 - p\left(\frac{1+f}{2+f}, N\right)$$
 and  $T$ .

Under the hypothesis that the second homolog is duplicated (i.e., an AB SNP becomes ABB, and a BA SNP becomes BAA), then  $\mathbf{R}_i$  has a Binomial distribution with parameters

$$1 - \frac{1+f}{2+f}$$

and T for AB SNPs, and

$$\frac{1+f}{2+f}$$

and T for BA SNPs. Therefore,

$$X_{i} = \begin{cases} 0 & wp1 - p\left(\frac{1+f}{2+f}, N_{i}\right) \\ 1 & wpp\left(\frac{1+f}{2+f}, N_{i}\right) \end{cases}$$

If we assume a constant depth of read N, this gives a Binomial distribution S with parameters

$$p\left(\frac{1+f}{2+f},N\right)$$
 and  $T$ .

Classification

As demonstrated in the sections above,  $X_i$  is a binary random variable with

$$Pr\{X_1=1\} = \begin{cases} p\left(\frac{1}{2}, N_i\right) & \text{given disomy} \\ p\left(\frac{1}{2-f}, N_i\right) & \text{homolog 1 deletion} \\ 1 - p\left(\frac{1}{2-f}, N_i\right) & \text{homolog 2 deletion} \\ 1 - p\left(\frac{1+f}{2+f}, N_i\right) & \text{homolog 1 duplication} \\ p\left(\frac{1+f}{2+f}, N_i\right) & \text{homolog 2 duplication} \end{cases}$$

This allows one to calculate the probability of the test statistic S under each hypothesis. The probability of each

72

hypothesis given the measured data can be calculated. In some embodiments, the hypothesis with the greatest probability is selected. If desired, the distribution on S can be simplified by either approximating each  $N_i$  with a constant depth of reach N or by truncating the depth of reads to a constant N. This simplification gives

10
$$S \sim \begin{cases} Bino\left(p\left(\frac{1}{2},N\right),T\right) & \text{given disomy} \\ Bino\left(p\left(\frac{1}{2-f},N\right),T\right) & \text{homolog 1 deletion} \end{cases}$$

$$S \sim \begin{cases} Bino\left(1-p\left(\frac{1}{2-f},N\right),T\right) & \text{homolog 2 deletion} \\ Bino\left(1-p\left(\frac{1+f}{2+f},N\right),T\right) & \text{homolog 1 duplication} \end{cases}$$

$$Bino\left(p\left(\frac{1+f}{2+f},N\right),T\right) & \text{homolog 2 duplication} \end{cases}$$

The value for f can be estimate by selecting the most likely value of f given the measured data, such as the value of f that generates the best data fit using an algorithm (e.g., a search algorithm) such as maximum likelihood estimation, maximum a-posteriori estimation, or Bayesian estimation. In some embodiments, multiple chromosome segments are analyzed and a value for f is estimated based on the data for each segment. If all the target cells have these duplications or deletions, the estimated values for f based on data for these different segments are similar. In some embodiments, f is experimentally measured such as by determining the fraction of DNA or RNA from cancer cells based on methylation differences (hypomethylation or hypermethylation) between cancer and non-cancerous DNA or RNA.

In some embodiments for mixed samples of fetal and maternal nucleic acids, the value of f is the fetal fraction, that 35 is the fraction of fetal DNA (or RNA) out of the total amount of DNA (or RNA) in the sample. In some embodiments, the fetal fraction is determined by obtaining genotypic data from a maternal blood sample (or fraction thereof) for a set of polymorphic loci on at least one chromosome that is 40 expected to be disomic in both the mother and the fetus; creating a plurality of hypotheses each corresponding to different possible fetal fractions at the chromosome; building a model for the expected allele measurements in the blood sample at the set of polymorphic loci on the chromosome for possible fetal fractions; calculating a relative probability of each of the fetal fractions hypotheses using the model and the allele measurements from the blood sample or fraction thereof; and determining the fetal fraction in the blood sample by selecting the fetal fraction corresponding to the hypothesis with the greatest probability. In some embodiments, the fetal fraction is determined by identifying those polymorphic loci where the mother is homozygous for a first allele at the polymorphic locus, and the father is (i) heterozygous for the first allele and a second 55 allele or (ii) homozygous for a second allele at the polymorphic locus; and using the amount of the second allele detected in the blood sample for each of the identified polymorphic loci to determine the fetal fraction in the blood sample (see, e.g., US Publ. No. 2012/0185176, filed Mar. 29, 60 2012, and US Pub. No. 2014/0065621, filed Mar. 13, 2013 which are each incorporated herein by reference in their entirety).

Another method for determining fetal fraction includes using a high throughput DNA sequencer to count alleles at a large number of polymorphic (such as SNP) genetic loci and modeling the likely fetal fraction (see, for example, US Publ. No. 2012/0264121, which is incorporated herein by

reference in its entirety). Another method for calculating fetal fraction can be found in Sparks et al.," Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18," Am J Obstet Gynecol 2012; 206:319.e1-9, 5 which is incorporated herein by reference in its entirety. In some embodiments, fetal fraction is determined using a methylation assay (see, e.g., U.S. Pat. Nos. 7,754,428; 7,901,884; and 8,166,382, which are each incorporated herein by reference in their entirety) that assumes certain loci are methylated or preferentially methylated in the fetus, and those same loci are unmethylated or preferentially unmethylated in the mother.

FIGS. 1A-13D are graphs showing the distribution of the test statistic S divided by T (the number of SNPs) ("S/T") for various copy number hypotheses for various depth of reads and tumor fractions (where f is the fraction of tumor DNA out of total DNA) for an increasing number of SNPs.

Single Hypothesis Rejection 20

The distribution of S for the disomy hypothesis does not depend on f. Thus, the probability of the measured data can be calculated for the disomy hypothesis without calculating f. A single hypothesis rejection test can be used for the null hypothesis of disomy. In some embodiments, the probability of S under the disomy hypothesis is calculated, and the hypothesis of disomy is rejected if the probability is below a given threshold value (such as less than 1 in 1,000). This indicates that a duplication or deletion of the chromosome segment is present. If desired, the false positive rate can be altered by adjusting the threshold value.

Exemplary Methods for Analysis of Phased Data Exemplary methods are described below for analysis of data from a sample known or suspected of being a mixed

sample containing DNA or RNA that originated from two or more cells that are not genetically identical. In some embodiments, phased data is used. In some embodiments, the method involves determining, for each calculated allele ratio, whether the calculated allele ratio is above or below the expected allele ratio and the magnitude of the difference for a particular locus. In some embodiments, a likelihood distribution is determined for the allele ratio at a locus for a particular hypothesis and the closer the calculated allele ratio is to the center of the likelihood distribution, the more 45 likely the hypothesis is correct. In some embodiments, the method involves determining the likelihood that a hypothesis is correct for each locus. In some embodiments, the method involves determining the likelihood that a hypothesis is correct for each locus, and combining the probabili- 50 ties of that hypothesis for each locus, and the hypothesis with the greatest combined probability is selected. In some embodiments, the method involves determining the likelihood that a hypothesis is correct for each locus and for each possible ratio of DNA or RNA from the one or more target 55 cells to the total DNA or RNA in the sample. In some embodiments, a combined probability for each hypothesis is determined by combining the probabilities of that hypothesis for each locus and each possible ratio, and the hypothesis with the greatest combined probability is selected.

In one embodiment, the following hypotheses are considered:  $H_{11}$  (all cells are normal),  $H_{10}$  (presence of cells with only homolog 1, hence homolog 2 deletion),  $H_{01}$  (presence of cells with only homolog 2, hence homolog 1 deletion),  $H_{21}$  (presence of cells with homolog 1 duplication),  $H_{12}$  (presence of cells with homolog 2 duplication). For a fraction f of target cells such as cancer cells or mosaic

74

cells (or the fraction of DNA or RNA from the target cells), the expected allele ratio for heterozygous (AB or BA) SNPs can be found as follows:

$$r(AB, H_{11}) = r(BA, H_{11}) = 0.5,$$
 Equation (1)  

$$r(AB, H_{10}) = r(BA, H_{01}) = \frac{1}{2 - f},$$
  

$$r(AB, H_{01}) = r(BA, H_{10}) = \frac{1 - f}{2 - f},$$
  

$$r(AB, H_{21}) = r(BA, H_{12}) = \frac{1 + f}{2 + f},$$
  

$$r(AB, H_{12}) = r(BA, H_{21}) = \frac{1}{2 + f}.$$

Bias, Contamination, and Sequencing Error Correction:

The observation Ds at the SNP consists of the number of original mapped reads with each allele present,  $n_A^{\ 0}$  and  $n_B^{\ 0}$ . Then, we can find the corrected reads  $n_A$  and  $n_B$  using the expected bias in the amplification of A and B alleles.

Let  $c_a$  to denote the ambient contamination (such as contamination from DNA in the air or environment) and  $r(c_a)$  to denote the allele ratio for the ambient contaminant (which is taken to be 0.5 initially). Moreover,  $c_g$  denotes the genotyped contamination rate (such as the contamination from another sample), and  $r(c_g)$  is the allele ratio for the contaminant. Let  $s_e(A,B)$  and  $s_e(B,A)$  denote the sequencing errors for calling one allele a different allele (such as by erroneously detecting an A allele when a B allele is present).

One can find the observed allele ratio  $q(r, c_a, r(c_a), c_g, r(c_g), s_e(A,B), s_e(B,A))$  for a given expected allele ratio r by correcting for ambient contamination, genotyped contamination, and sequencing error.

Since the contaminant genotypes are unknown, population frequencies can be used to find  $P(r(c_g))$ . More specifically, let p be the population frequency for one of the alleles (which may be referred to as a reference allele). Then, we have  $P(r(c_g)=0)=(1-p)^2$ ,  $P(r(c_g)=0)=2p(1-p)$ , and  $P(r(cg)=0)=p^2$ . The conditional expectation over  $r(c_g)$  can be used to determine the  $E[q(r, c_a, r(c_a), c_g, r(c_g), s_e(A,B), s_e(B,A))]$ . Note that the ambient and genotyped contamination are determined using the homozygous SNPs, hence they are not affected by the absence or presence of deletions or duplications. Moreover, it is possible to measure the ambient and genotyped contamination using a reference chromosome if desired.

Likelihood at Each SNP:

The equation below gives the probability of observing  $n_A$  and  $n_B$  given an allele ratio r:

$$P(n_A, n_B \mid r) = p_{bino}(n_A; n_A + n_B, r) = \binom{n_A + n_B}{n_A} r^{n_A} (1 - r)^{n_B}.$$
 Equation (2)

Let  $D_s$  denote the data for SNP s. For each hypothesis  $h \in \{H_{11}, H_{01}, H_{10}, H_{21}, H_{12}\}$ , one can let r = r(AB,h) or r = r(BA,h) in the equation (1) and find the conditional expectation over  $r(c_g)$  to determine the observed allele ratio  $E[q(r, c_a, r(c_a), c_g, r(c_g))]$ . Then, letting  $r = E[q(r, c_a, r(c_a), c_g, r(c_g), s_e(A,B), s_e(B,A))]$  in equation (2) one can determine  $P(D_s|h,f)$ .

Search Algorithm:

In some embodiments, SNPs with allele ratios that seem to be outliers are ignored (such as by ignoring or eliminating SNPs with allele ratios that are at least 2 or 3 standard

deviations above or below the mean value). Note that an advantage identified for this approach is that in the presence of higher mosaicism percentage, the variability in the allele ratios may be high, hence this ensures that SNPs will not be trimmed due to mosaicism.

Let  $F = \{f_1, \ldots, f_N\}$  denote the search space for the mosaicism percentage (such as the tumor fraction). One can determine  $P(D_s|h,f)$  at each SNP s and  $f \in F$ , and combine the likelihood over all SNPs.

The algorithm goes over each f for each hypothesis. Using 10 a search method, one concludes that mosaicism exists if there is a range F\* of f where the confidence of the deletion or duplication hypothesis is higher than the confidence of the no deletion and no duplication hypotheses. In some embodiments, the maximum likelihood estimate for  $P(D_s|h,f)$  in F\* 15 is determined. If desired, the confidence for each hypothesis can be determined.

#### Additional Embodiments

In some embodiments, a beta binomial distribution is used instead of binomial distribution. In some embodiments, a reference chromosome or chromosome segment is used to determine the sample specific parameters of beta binomial. 25 Theoretical Performance Using Simulations:

If desired, one can evaluate the theoretical performance of the algorithm by randomly assigning number of reference reads to a SNP with given depth of read (DOR). For the normal case, use p=0.5 for the binomial probability parameter, and for deletions or duplications, p is revised accordingly. Exemplary input parameters for each simulation are as follows: (1) number of SNPs S (2) constant DOR D per SNP, (3) p, and (4) number of experiments.

First Simulation Experiment: 35 This experiment focused on S $\in$ {500, 1000}, D $\in$ {500, 1000} and p $\in$ {0%, 1%, 2%, 3%, 4%, 5%}. We performed 1,000 simulation experiments in each setting (hence 24,000 experiments with phase, and 24,000 without phase). We simulated the number of reads from a binomial distribution 40 (if desired, other distributions can be used). The false positive rate (in the case of p>0%) were determined both with or without phase information. False positive rates are listed in FIG. 26. Note that phase information is very helpful, especially for 45 S=1000, D=1000. Although for S=500, D=500, the algorithm has the highest false positive rates with or without phase out of the conditions tested. False negative rates are listed in FIG. 27.

Phase information is particularly useful for low mosaicism percentages ( $\leq 3\%$ ). Without phase information, a high level of false negatives were observed for p=1% because the confidence on deletion is determined by assigning equal chance to  $H_{10}$  and  $H_{01}$ , and a small deviation in favor of one hypothesis is not sufficient to compensate for the low 55 likelihood from the other hypothesis. This applies to duplications as well. Note also that the algorithm seems to be more sensitive to depth of read compared to number of SNPs. For the results with phase information, we assume that perfect phase information is available for a high number 60 of consecutive heterozygous SNPs. If desired, haplotype information can be obtained by probabilistically combining haplotypes on smaller segments.

Second Simulation Experiment:

This experiment focused on Se{100, 200, 300, 400, 500}, 65 De{1000, 2000, 3000, 4000, 5000} and pe{0%, 1%, 1.5%, 2%, 2.5%, 3%} and 10000 random experiments at each

76

setting. The false positive rate (in the case of p=0%) and false negative rate (in the case of p>0%) were determined both with or without phase information. The false negative rate is below 10% for D≥3000 and N≥200 using haplotype information, whereas the same performance is reached for D=5000 and N≥400 (FIGS. **20**A and **20**B). The difference between the false negative rate was particularly stark for small mosaicism percentages (FIGS. **21**A-**25**B). For example, when p=1%, a less than 20% false negative rate is never reached without haplotype data, whereas it is close to 0% for N≥300 and D≥3000. For p=3%, a 0% false negative rate is observed with haplotype data, while N≥300 and D≥3000 is needed to reach the same performance without haplotype data

Exemplary Methods for Detecting Deletions and Duplications without Phased Data

In some embodiments, unphased genetic data is used to determine if there is an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of an individual (such as in the genome of one or more cells or in cfDNA or cfRNA). In some embodiments, phased genetic data is used but the phasing is ignored. In some embodiments, the sample of DNA or RNA is a mixed sample of cfDNA or cfRNA from the individual that includes cfDNA or cfRNA from two or more genetically different cells. In some embodiments, the method utilizes the magnitude of the difference between the calculated allele ratio and the expected allele ratio for each of the loci.

In some embodiments, the method involves obtaining genetic data at a set of polymorphic loci on the chromosome or chromosome segment in a sample of DNA or RNA from one or more cells from the individual by measuring the quantity of each allele at each locus. In some embodiments, allele ratios are calculated for the loci that are heterozygous in at least one cell from which the sample was derived (such as the loci that are heterozygous in the fetus and/or heterozygous in the mother). In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles divided by the total measured quantity of all the alleles for the locus. In some embodiments, the calculated allele ratio for a particular locus is the measured quantity of one of the alleles (such as the allele on the first homologous chromosome segment) divided by the measured quantity of one or more other alleles (such as the allele on the second homologous chromosome segment) for the locus. The calculated allele ratios and expected allele ratios may be calculated using any of the methods described herein or any standard method (such as any mathematical transformation of the calculated allele ratios or expected allele ratios described herein).

In some embodiments, a test statistic is calculated based on the magnitude of the difference between the calculated allele ratio and the expected allele ratio for each of the loci. In some embodiments, the test statistic  $\Delta$  is calculated using the following formula

$$\Delta = \frac{\Sigma_{AllLoci}(\delta_i - \mu_i)}{\sqrt{\Sigma_{AllLoci}\sigma_i^2}}$$

wherein  $\delta_i$  is the magnitude of the difference between the calculated allele ratio and the expected allele ratio for the ith loci:

wherein  $\mu_i$  is the mean value of  $\delta_i$ ; and wherein  $\sigma_i^2$  is the standard deviation of  $\delta_i$ .

# US 11,530,454 B2

7

For example, we can define  $\delta_i$  as follows when the expected allele ratio is 0.5:

 $\delta_i \triangleq |1/2 - R_i|$ .

Values for  $\mu_i$  and  $\sigma_i$  can be computed using the fact that  $R_i$  is a Binomial random variable. In some embodiments, the standard deviation is assumed to be the same for all the loci. In some embodiments, the average or weighted average value of the standard deviation or an estimate of the standard deviation is used for the value of  $\sigma_i^2$ . In some embodiments, the test statistic is assumed to have a normal distribution. For example, the central limit theorem implies that the distribution of  $\Delta$  converges to a standard normal as the number of loci (such as the number of SNPs T) grows large.

In some embodiments, a set of one or more hypotheses 15 specifying the number of copies of the chromosome or chromosome segment in the genome of one or more of the cells are enumerated. In some embodiments, the hypothesis that is most likely based on the test statistic is selected, thereby determining the number of copies of the chromo- 20 some or chromosome segment in the genome of one or more of the cells. In some embodiments, a hypotheses is selected if the probability that the test statistic belongs to a distribution of the test statistic for that hypothesis is above an upper threshold; one or more of the hypotheses is rejected if the 25 probability that the test statistic belongs to the distribution of the test statistic for that hypothesis is below an lower threshold; or a hypothesis is neither selected nor rejected if the probability that the test statistic belongs to the distribution of the test statistic for that hypothesis is between the 30 lower threshold and the upper threshold, or if the probability is not determined with sufficiently high confidence. In some embodiments, an upper and/or lower threshold is determined from an empirical distribution, such as a distribution from training data (such as samples with a known copy number, 3: such as diploid samples or samples known to have a particular deletion or duplication). Such an empirical distribution can be used to select a threshold for a single hypothesis rejection test.

Note that the test statistic  $\Delta$  is independent of S and 40 therefore both can be used independently, if desired. Exemplary Methods for Detecting Deletions and Duplications Using Allele Distributions or Patterns

This section includes methods for determining if there is an overrepresentation of the number of copies of a first 45 homologous chromosome segment as compared to a second homologous chromosome segment. In some embodiments, the method involves enumerating (i) a plurality of hypotheses specifying the number of copies of the chromosome or chromosome segment that are present in the genome of one 50 or more cells (such as cancer cells) of the individual or (ii) a plurality of hypotheses specifying the degree of overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of one or more cells of 55 the individual. In some embodiments, the method involves obtaining genetic data from the individual at a plurality of polymorphic loci (such as SNP loci) on the chromosome or chromosome segment. In some embodiments, a probability distribution of the expected genotypes of the individual for 60 each of the hypotheses is created. In some embodiments, a data fit between the obtained genetic data of the individual and the probability distribution of the expected genotypes of the individual is calculated. In some embodiments, one or more hypotheses are ranked according to the data fit, and the 65 hypothesis that is ranked the highest is selected. In some embodiments, a technique or algorithm, such as a search

78

algorithm, is used for one or more of the following steps: calculating the data fit, ranking the hypotheses, or selecting the hypothesis that is ranked the highest. In some embodiments, the data fit is a fit to a beta-binomial distribution or a fit to a binomial distribution. In some embodiments, the technique or algorithm is selected from the group consisting of maximum likelihood estimation, maximum a-posteriori estimation, Bayesian estimation, dynamic estimation (such as dynamic Bayesian estimation), and expectation-maximization estimation. In some embodiments, the method includes applying the technique or algorithm to the obtained genetic data and the expected genetic data.

In some embodiments, the method involves enumerating (i) a plurality of hypotheses specifying the number of copies of the chromosome or chromosome segment that are present in the genome of one or more cells (such as cancer cells) of the individual or (ii) a plurality of hypotheses specifying the degree of overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome of one or more cells of the individual. In some embodiments, the method involves obtaining genetic data from the individual at a plurality of polymorphic loci (such as SNP loci) on the chromosome or chromosome segment. In some embodiments, the genetic data includes allele counts for the plurality of polymorphic loci. In some embodiments, a joint distribution model is created for the expected allele counts at the plurality of polymorphic loci on the chromosome or chromosome segment for each hypothesis. In some embodiments, a relative probability for one or more of the hypotheses is determined using the joint distribution model and the allele counts measured on the sample, and the hypothesis with the greatest probability is selected.

In some embodiments, the distribution or pattern of alleles (such as the pattern of calculated allele ratios) is used to determine the presence or absence of a CNV, such as a deletion or duplication. If desired the parental origin of the CNV can be determined based on this pattern. A maternally inherited duplication is an extra copy of a chromosome segment from the mother, and maternally inherited deletion is the absence of the copy of a chromosome segment from the mother such that the only copy of the chromosome segment that is present is from the father. Exemplary patterns are illustrated in FIGS. **15**A-**19**D and are described further below.

To determine the presence or absence of a deletion of a chromosome segment of interest, the algorithm considers the distribution of sequence counts from each of two possible alleles at large number of SNPs per chromosome. It is important to note that some embodiments of the algorithm use an approach that does not lend itself to visualization. Thus, for the purposes of illustration, the data is displayed in FIGS. 15A-18 in a simplified fashion as ratios of the two most likely alleles, labeled as A and B, so that the relevant trends can be more readily visualized. This simplified illustration does not take into account some of the possible features of the algorithm. For example, two embodiments for the algorithm that are not possible to illustrate with a method of visualization that displays allele ratios are: 1) the ability to leverage linkage disequilibrium, i.e. the influence that a measurement at one SNP has on the likely identity of a neighboring SNP, and 2) the use of non-Gaussian data models that describe the expected distribution of allele measurements at a SNP given platform characteristics and amplification biases. Also note that a simplified version of the algorithm only considers the two most common alleles at each SNP, ignoring other possible alleles.

# US 11,530,454 B2

79

Deletions of interest were detected in genomic and maternal blood samples. In some embodiments, the genomic and maternal plasma samples are analyzed using the multiplex-PCR and sequencing method of Example 1. The genomic DNA syndrome samples tested lacked heterozygous SNPs in the targeted regions, confirming the ability of the assays to distinguish monosomy (affected) from disomy (unaffected). Analysis of cfDNA from a maternal blood sample was able to detect 22q11.2 deletion syndrome, Cri-du-Chat deletion syndrome, and Wolf-Hirschhorn deletion syndrome, as well as the other deletion syndromes in FIG. 14 in the fetus.

FIGS. 15A-15C depict data that indicate the presence of two chromosomes when the sample is entirely maternal (no fetal cfDNA present, FIG. 15A), contains a moderate fetal cfDNA fraction of 12% (FIG. 15B), or contains a high fetal cfDNA fraction of 26% (FIG. 15C). The x-axis represents the linear position of the individual polymorphic loci along the chromosome, and the y-axis represents the number of A allele reads as a fraction of the total (A+B) allele reads. 20 Maternal and fetal genotypes are indicated to the right of the plots. The plots are color-coded according to maternal genotype, such that red indicates a maternal genotype of AA, blue indicates a maternal genotype of BB, and green indicates a maternal genotype of AB. Note that the measure- 25 ments are made on total cfDNA isolated from maternal blood, and the cfDNA includes both maternal and fetal cfDNA; thus, each spot represents the combination of the fetal and maternal DNA contribution for that SNP. Therefore, increasing the proportion of maternal cfDNA from 0% 30 to 100% will gradually shift some spots up or down within the plots, depending on the maternal and fetal genotype.

In all cases, SNPs that are homozygous for the A allele (AA) in both the mother and the fetus are found tightly associated with the upper limit of the plots, as the fraction 35 of A allele reads is high because there should be no B alleles present. Conversely, SNPs that are homozygous for the B allele in both the mother and the fetus are found tightly associated with the lower limit of the plots, as the fraction of A allele reads is low because there should be only B 40 alleles. The spots that are not tightly associated with the upper and lower limits of the plots represent SNPs for which the mother, the fetus, or both are heterozygous; these spots are useful for identifying fetal deletions or duplications, but can also be informative for determining paternal versus 4. maternal inheritance. These spots segregate based on both maternal and fetal genotypes and fetal fraction, and as such the precise position of each individual spot along the y-axis depends on both stoichiometry and fetal fraction. For example, loci where the mother is AA and the fetus is AB are 50 expected to have a different fraction of A allele reads, and thus different positioning along the y-axis, depending on the fetal fraction.

FIG. 15A has data for a non-pregnant woman, and thus represents the pattern when the genotype is entirely maternal. This pattern includes "clusters" of spots: a red cluster tightly associated with the top of the plot (SNPs where the maternal genotype is AA), a blue cluster tightly associated with the bottom of the plot (SNPs where the maternal genotype is BB), and a single, centered green cluster (SNPs where the maternal genotype is AB). For FIG. 15B, the contribution of fetal alleles to the fraction of A allele reads shifts the position of some allele spots up or down along the y-axis. For FIG. 15C, the pattern, including two red and two blue peripheral bands and a trio of central green bands, is readily apparent. The three central green bands correspond to SNPs that are heterozygous in the mother, and two

80

"peripheral" bands each at both the top (red) and bottom (blue) of the plots correspond to SNPs that are homozygous in the mother.

Analysis of a 22q11.2 deletion carrier (a mother with this deletion) is shown in FIG. 16A. The deletion carrier does not have heterozygous SNPs in this region since the carrier only has one copy of this region. Thus, this deletion is indicated by the absence of the green AB SNPs. The analysis of a paternally inherited 22q11 deletion in a fetus is shown in FIG. 16B. When the fetus only inherits a single copy of a chromosome segment (in the case of a paternally inherited deletion, the copy present in the fetus comes from the mother), and thus only inherits a single allele for each locus in this segment, heterozygosity of the fetus is not possible. As such, the only possible fetal SNP identities are A or B. Note the absence of internal peripheral bands. For a paternally inherited deletion, the characteristic pattern includes two central green bands that represent SNPs for which the mother is heterozygous, and only has single peripheral red and blue bands that represent SNPs for which the mother is homozygous, and which remain tightly associated with the upper and lower limits of the plots (1 and 0), respectively.

Analysis of a maternally inherited Cri-du-Chat deletion syndrome is shown in FIG. 17. There are two central green bands instead of three green bands, and there are two red and two blue peripheral bands. A maternally inherited deletion (such as a maternal carrier of Duchenne's muscular dystrophy) can also be detected based on the small amount of signal in that region of the deletion in a mixed sample of maternal and fetal DNA (such as a plasma sample) due to both the mother and the fetus having the deletion.

FIG. 18 is a plot of a paternally inherited Wolf-Hirschhorn deletion syndrome, as indicated by the presence of one red and one blue peripheral band.

If desired, similar plots can be generated for a sample from an individual suspected of having a deletion or duplication, such as a CNV associated with cancer. In such plots, the following color coding can be used based on the genotype of cells without the CNV: red indicates a genotype of AA, blue indicates a genotype of BB, and green indicates a genotype of AB. In some embodiments for a deletion, the pattern includes two central green bands that represent SNPs for which the individual is heterozygous (top green band represents AB from cells without the deletion and A from cells with the deletion, and bottom green band represents AB from cells without the deletion and B from cells with the deletion), and only has single peripheral red and blue bands that represent SNPs for which the individual is homozygous, and which remain tightly associated with the upper and lower limits of the plots (1 and 0), respectively. In some embodiments, the separation of the two green bands increases as the fraction of cells, DNA, or RNA with the deletion increases

Exemplary Methods for Identifying and Analyzing Multiple Pregnancies

In some embodiments, any of the methods of the present invention are used to detect the presence of a multiple pregnancy, such as a twin pregnancy, where at least one of the fetuses is genetically different from at least one other fetus. In some embodiments, fraternal twins are identified based on the presence of two fetus with different allele, different allele ratios, or different allele distributions at some (or all) of the tested loci. In some embodiments, fraternal twins are identified by determining the expected allele ratio at each locus (such as SNP loci) for two fetuses that may have the same or different fetal fractions in the sample (such as a plasma sample). In some embodiments, the likelihood

# US 11,530,454 B2

of a particular pair of fetal fractions (where fl is the fetal fraction for fetus 1, and f2 is the fetal fraction for fetus 2) is calculated by considering some or all of the possible genotypes of the two fetuses, conditioned on the mother's genotype and genotype population frequencies. The mixture of two fetal and one maternal genotype, combined with the fetal fractions, determine the expected allele ratio at a SNP. For example, if the mother is AA, fetus 1 is AA, and fetus 2 is AB, the overall fraction of B allele at the SNP is one-half of f2. The likelihood calculation asks how well all of the 10 SNPs together match the expected allele ratios based on all of the possible combinations of fetal genotypes. The fetal fraction pair (f1, f2) that best matches the data is selected. It is not necessary to calculated specific genotypes of the fetuses; instead, one can, for example, considered all of the 15 possible genotypes in a statistical combination. In some embodiments, if the method does not distinguish between singleton and identical twins, an ultrasound can be performed to determine whether there is a singleton or identical twin pregnancy. If the ultrasound detects a twin pregnancy 20 it can be assumed that the pregnancy is an identical twin pregnancy because a fraternal twin pregnancy would have been detected based on the SNP analysis discussed above.

In some embodiments, a pregnant mother is known to have a multiple pregnancy (such as a twin pregnancy) based 25 on prior testing, such as an ultrasound. Any of the methods of the present invention can be used to determine whether the multiple pregnancy includes identical or fraternal twins. For example, the measured allele ratios can be compared to what would be expected for identical twins (the same allele 30 ratios as a singleton pregnancy) or for fraternal twins (such as the calculation of allele ratios as described above). Some identical twins are monochorionic twins, which have a risk of twin-to-twin transfusion syndrome. Thus, twins determined to be identical twins using a method of the invention 3: are desirably tested (such as by ultrasound) to determine if they are monochorionic twins, and if so, these twins can be monitored (such as bi-weekly ultrasounds from 16 weeks) for signs of win-to-twin transfusion syndrome.

In some embodiments, any of the methods of the present 40 invention are used to determine whether any of the fetuses in a multiple pregnancy, such as a twin pregnancy, are aneuploid. Aneuploidy testing for twins begins with the fetal fraction estimate. In some embodiments, the fetal fraction pair (f1, f2) that best matches the data is selected as 45 described above. In some embodiments, a maximum likelihood estimate is performed for the parameter pair (f1, f2) over the range of possible fetal fractions. In some embodiments, the range of f2 is from 0 to f1 because f2 is defined as the smaller fetal fraction. Given a pair (f1, f2), data 50 likelihood is calculated from the allele ratios observed at a set of loci such as SNP loci. In some embodiments, the data likelihood reflects the genotypes of the mother, the father if available, population frequencies, and the resulting probabilities of fetal genotypes. In some embodiments, SNPs are 55 assumed independent. The estimated fetal fraction pair is the one that produces the highest data likelihood. If f2 is 0 then the data is best explained by only one set of fetal genotypes, indicating identical twins, where f1 is the combined fetal fraction. Otherwise f1 and f2 are the estimates of the 60 individual twin fetal fractions. Having established the best estimate of (f1, f2), one can predict the overall fraction of B allele in the plasma for any combination of maternal and fetal genotypes, if desired. It is not necessary to assign individual sequence reads to the individual fetuses. Ploidy 65 testing is performed using another maximum likelihood estimate which compares the data likelihood of two hypoth-

eses. In some embodiments for identical twins, one consider the hypotheses (i) both twins are euploid, and (ii) both twins are trisomic. In some embodiments for fraternal twins, one considers the hypotheses (i) both twins are euploid and (ii) at least one twin is trisomic. The trisomy hypotheses for fraternal twins are based on the lower fetal fraction, since a trisomy in the twin with a higher fetal fraction would also be detected. Ploidy likelihoods are calculated using a method which predicts the expected number of reads at each targeted genome locus conditioned on either the disomy or trisomy hypothesis. There is no requirement for a disomy reference chromosome. The variance model for the expected number of reads takes into account the performance of individual target loci as well as the correlation between loci (see, for example, U.S. Ser. No. 62/008,235, filed Jun. 5, 2014, and U.S. Ser. No. 62/032,785, filed Aug. 4, 2014, which are each hereby incorporated by reference in its entirety). If the smaller twin has fetal fraction f1, our ability to detect a trisomy in that twin is equivalent to our ability to detect a trisomy in a singleton pregnancy at the same fetal fraction. This is because the part of the method that detects the trisomy in some embodiments does not depend on genotypes and does not distinguish between multiple or singleton pregnancy. It simply looks for an increased number of reads in accordance with the determined fetal fraction.

In some embodiments, the method includes detecting the presence of twins based on SNP loci (such as described above). If twins are detected, SPNs are used to determine the fetal fraction of each fetus (f1, f2) such as described above. In some embodiments, samples that have high confidence disomy calls are used to determine the amplification bias on a per-SNP basis. In some embodiments, these samples with high confidence disomy calls are analyzed in the same run as one or more samples of interest. In some embodiments, the amplification bias on a per-SNP basis is used to model the distribution of reads for one or more chromosomes or chromosome segments of interest such as chromosome 21 that are expected or the disomy hypothesis and the trisomy hypothesis given the lower of the two twin fetal fraction. The likelihood or probability of disomy or trisomy is calculated given the two models and the measured quantity of the chromosome or chromosome segment of interest.

In some embodiments, the threshold for a positive aneuploidy call (such as a trisomy call) is set based on the twin with the lower fetal fraction. This way, if the other twin is positive, or if both are positive, the total chromosome representation is definitely above the threshold.

Exemplary Counting Methods/Quantitative Methods In some embodiments, one or more counting methods (also referred to as quantitative methods) are used to detect one or more CNS, such as deletions or duplications of chromosome segments or entire chromosomes. In some embodiments, one or more counting methods are used to determine whether the overrepresentation of the number of copies of the first homologous chromosome segment is due to a duplication of the first homologous chromosome segment or a deletion of the second homologous chromosome segment. In some embodiments, one or more counting methods are used to determine the number of extra copies of a chromosome segment or chromosome that is duplicated (such as whether there are 1, 2, 3, 4, or more extra copies). In some embodiments, one or more counting methods are used to differentiate a sample has many duplications and a smaller tumor fraction from a sample with fewer duplications and a larger tumor fraction. For example, one or more counting methods may be used to differentiate a sample with four extra chromosome copies and a tumor fraction of 10%

# US 11,530,454 B2

83

from a sample with two extra chromosome copies and a tumor fraction of 20%. Exemplary methods are disclosed, e.g. U.S. Publication Nos. 2007/0184467; 2013/0172211; and 2012/0003637; U.S. Pat. Nos. 8,467,976; 7,888,017; 8,008,018; 8,296,076; and 8,195,415; U.S. Ser. No. 62/008, 235, filed Jun. 5, 2014, and U.S. Ser. No. 62/032,785, filed Aug. 4, 2014, which are each hereby incorporated by reference in its entirety.

In some embodiment, the counting method includes counting the number of DNA sequence-based reads that map to one or more given chromosomes or chromosome segments. Some such methods involve creation of a reference value (cut-off value) for the number of DNA sequence reads mapping to a specific chromosome or chromosome segment, wherein a number of reads in excess of the value is indicative of a specific genetic abnormality.

In some embodiments, the total measured quantity of all the alleles for one or more loci (such as the total amount of a polymorphic or non-polymorphic locus) is compared to a 20 reference amount. In some embodiments, the reference amount is (i) a threshold value or (ii) an expected amount for a particular copy number hypothesis. In some embodiments, the reference amount (for the absence of a CNV) is the total measured quantity of all the alleles for one or more loci for  $^{\,\,25}$ one or more chromosomes or chromosomes segments known or expected to not have a deletion or duplication. In some embodiments, the reference amount (for the presence of a CNV) is the total measured quantity of all the alleles for one or more loci for one or more chromosomes or chromosomes segments known or expected to have a deletion or duplication. In some embodiments, the reference amount is the total measured quantity of all the alleles for one or more loci for one or more reference chromosomes or chromosome 35 segments. In some embodiments, the reference amount is the mean or median of the values determined for two or more different chromosomes, chromosome segments, or different samples. In some embodiments, random (e.g., massively parallel shotgun sequencing) or targeted sequencing is used 40 to determine the amount of one or more polymorphic or non-polymorphic loci.

In some embodiments utilizing a reference amount, the method includes (a) measuring the amount of genetic material on a chromosome or chromosome segment of interest; 45 (b) comparing the amount from step (a) to a reference amount; and (c) identifying the presence or absence of a deletion or duplication based on the comparison.

In some embodiments utilizing a reference chromosome or chromosome segment, the method includes sequencing 50 DNA or RNA from a sample to obtain a plurality of sequence tags aligning to target loci. In some embodiments, the sequence tags are of sufficient length to be assigned to a specific target locus (e.g., 15-100 nucleotides in length); the target loci are from a plurality of different chromosomes or 55 chromosome segments that include at least one first chromosome or chromosome segment suspected of having an abnormal distribution in the sample and at least one second chromosome or chromosome segment presumed to be normally distributed in the sample. In some embodiments, the 60 plurality of sequence tags are assigned to their corresponding target loci. In some embodiments, the number of sequence tags aligning to the target loci of the first chromosome or chromosome segment and the number of sequence tags aligning to the target loci of the second 65 chromosome or chromosome segment are determined. In some embodiments, these numbers are compared to deter84

mine the presence or absence of an abnormal distribution (such as a deletion or duplication) of the first chromosome or chromosome segment.

In some embodiments, the value of f (such as the fetal fraction or tumor fraction) is used in the CNV determination, such as to compare the observed difference between the amount of two chromosomes or chromosome segments to the difference that would be expected for a particular type of CNV given the value of f (see, e.g., US Publication No 2012/0190020; US Publication No 2012/0190021; US Publication No 2012/0190557; US Publication No 2012/ 0191358, which are each hereby incorporated by reference in its entirety). For example, the difference in the amount of a chromosome segment that is duplicated in a fetus compared to a disomic reference chromosome segment in a blood sample from a mother carrying the fetus increases as the fetal fraction increases. Additionally, the difference in the amount of a chromosome segment that is duplicated in a tumor compared to a disomic reference chromosome segment increases as the tumor fraction increases. In some embodiments, the method includes comparing the relative frequency of a chromosome or chromosome segment of interest to a reference chromosomes or chromosome segment (such as a chromosome or chromosome segment expected or known to be disomic) to the value of f to determine the likelihood of the CNV. For example, the difference in amounts between the first chromosomes or chromosome segment to the reference chromosome or chromosome segment can be compared to what would be expected given the value of f for various possible CNVs (such as one or two extra copies of a chromosome segment of interest).

The following prophetic examples illustrate the use of a counting method/quantitative method to differentiate between a duplication of the first homologous chromosome segment and a deletion of the second homologous chromosome segment. If one considers the normal disomic genome of the host to be the baseline, then analysis of a mixture of normal and cancer cells yields the average difference between the baseline and the cancer DNA in the mixture. For example, imagine a case where 10% of the DNA in the sample originated from cells with a deletion over a region of a chromosome that is targeted by the assay. In some embodiments, a quantitative approach shows that the quantity of reads corresponding to that region is expected to be 95% of what is expected for a normal sample. This is because one of the two target chromosomal regions in each of the tumor cells with a deletion of the targeted region is missing, and thus the total amount of DNA mapping to that region is 90%(for the normal cells) plus ½×10% (for the tumor cells) =95%. Alternately in some embodiments, an allelic approach shows that the ratio of alleles at heterozygous loci averaged 19:20. Now imagine a case where 10% of the DNA in the sample originated from cells with a five-fold focal amplification of a region of a chromosome that is targeted by the assay. In some embodiments, a quantitative approach shows that the quantity of reads corresponding to that region is expected to be 125% of what is expected for a normal sample. This is because one of the two target chromosomal regions in each of the tumor cells with a five-fold focal amplification is copied an extra five times over the targeted region, and thus the total amount of DNA mapping to that region is 90% (for the normal cells) plus (2+5)×10%/2 (for the tumor cells)=125%. Alternately in some embodiments, an allelic approach shows that the ratio of alleles at heterozygous loci averaged 25:20. Note that when using an allelic approach alone, a focal amplification of five-fold over

US 11,530,454 B2

85

a chromosomal region in a sample with 10% cfDNA may appear the same as a deletion over the same region in a sample with 40% cfDNA; in these two cases, the haplotype that is under-represented in the case of the deletion appears to be the haplotype without a CNV in the case with the focal duplication, and the haplotype without a CNV in the case of the deletion appears to be the over-represented haplotype in the case with the focal duplication. Combining the likelihoods produced by this allelic approach with likelihoods produced by a quantitative approach differentiates between 10 the two possibilities.

Exemplary Counting Methods/Quantitative Methods Using Reference Samples

An exemplary quantitative method that uses one or more reference samples is described in U.S. Ser. No. 62/008,235, 15 filed Jun. 5, 2014 and U.S. Ser. No. 62/032,785, filed Aug. 4, 2014, which is hereby incorporated by reference in its entirety. In some embodiments, one or more reference samples most likely to not have any CNVs on one or more chromosomes or chromosomes of interest (e.g., a normal 20 sample) are identified by selecting the samples with the highest fraction of tumor DNA, selecting the samples with the z-score closest to zero, selecting the samples where the data fits the hypothesis corresponding to no CNVs with the highest confidence or likelihood, selecting the samples 25 known to be normal, selecting the samples from individuals with the lowest likelihood of having cancer (e.g., having a low age, being a male when screening for breast cancer, having no family history, etc.), selecting the samples with the highest input amount of DNA, selecting the samples with 30 the highest signal to noise ratio, selecting samples based on other criteria believed to be correlated to the likelihood of having cancer, or selecting samples using some combination of criteria. Once the reference set is chosen, one can make the assumption that these cases are disomic, and then 3: estimate the per-SNP bias, that is, the experiment-specific amplification and other processing bias for each locus. Then, one can use this experiment-specific bias estimate to correct the bias in the measurements of the chromosome of interest. such as chromosome 21 loci, and for the other chromosome 40 loci as appropriate, for the samples that are not part of the subset where disomy is assumed for chromosome 21. Once the biases have been corrected for in these samples of unknown ploidy, the data for these samples can then be analyzed a second time using the same or a different method 45 to determine whether the individuals (such as fetuses) are afflicted with trisomy 21. For example, a quantitative method can be used on the remaining samples of unknown ploidy, and a z-score can be calculated using the corrected measured genetic data on chromosome 21. Alternately, as 50 part of the preliminary estimate of the ploidy state of chromosome 21, a fetal fraction (or tumor fraction for samples from an individual suspected of having cancer) can be calculated. The proportion of corrected reads that are expected in the case of a disomy (the disomy hypothesis), 55 and the proportion of corrected reads that are expected in the case of a trisomy (the trisomy hypothesis) can be calculated for a case with that fetal fraction. Alternately, if the fetal fraction was not measured previously, a set of disomy and trisomy hypotheses can be generated for different fetal 60 fractions. For each case, an expected distribution of the proportion of corrected reads can be calculated given expected statistical variation in the selection and measurement of the various DNA loci. The observed corrected proportion of reads can be compared to the distribution of 65 the expected proportion of corrected reads, and a likelihood ratio can be calculated for the disomy and trisomy hypoth80

eses, for each of the samples of unknown ploidy. The ploidy state associated with the hypothesis with the highest calculated likelihood can be selected as the correct ploidy state.

In some embodiments, a subset of the samples with a sufficiently low likelihood of having cancer may be selected to act as a control set of samples. The subset can be a fixed number, or it can be a variable number that is based on choosing only those samples that fall below a threshold. The quantitative data from the subset of samples may be combined, averaged, or combined using a weighted average where the weighting is based on the likelihood of the sample being normal. The quantitative data may be used to determine the per-locus bias for the amplification the sequencing of samples in the instant batch of control samples. The per-locus bias may also include data from other batches of samples. The per-locus bias may indicate the relative overor under-amplification that is observed for that locus compared to other loci, making the assumption that the subset of samples do not contain any CNVs, and that any observed over or under-amplification is due to amplification and/or sequencing or other bias. The per-locus bias may take into account the GC content of the amplicon. The loci may be grouped into groups of loci for the purpose of calculating a per-locus bias. Once the per-locus bias has been calculated for each locus in the plurality of loci, the sequencing data for one or more of the samples that are not in the subset of the samples, and optionally one or more of the samples that are in the subset of samples, may be corrected by adjusting the quantitative measurements for each locus to remove the effect of the bias at that locus. For example, if SNP 1 was observed, in the subset of patients, to have a depth of read that is twice as great as the average, the adjustment may involve replacing the number of reads corresponding from SNP 1 with a number that is half as great. If the locus in question is a SNP, the adjustment may involve cutting the number of reads corresponding to each of the alleles at that locus in half. Once the sequencing data for each of the loci in one or more samples has been adjusted, it may be analyzed using a method for the purpose of detecting the presence of a CNV at one or more chromosomal regions.

In an example, sample A is a mixture of amplified DNA originating from a mixture of normal and cancerous cells that is analyzed using a quantitative method. The following illustrates exemplary possible data. A region of the q arm on chromosome 22 is found to only have 90% as much DNA mapping to that region as expected; a focal region corresponding to the HER2 gene is found to have 150% as much DNA mapping to that region as expected; and the p-arm of chromosome 5 is found to have 105% as much DNA mapping to it as expected. A clinician may infer that the sample has a deletion of a region on the q arm on chromosome 22, and a duplication of the HER2 gene. The clinician may infer that since the 22q deletions are common in breast cancer, and that since cells with a deletion of the 22q region on both chromosomes usually do not survive, that approximately 20% of the DNA in the sample came from cells with a 22q deletion on one of the two chromosomes. The clinician may also infer that if the DNA from the mixed sample that originated from tumor cells originated from a set of genetically tumor cells whose HER2 region and 22q regions were homogenous, then the cells contained a five-fold duplication of the HER2 region.

In an example, Sample A is also analyzed using an allelic method. The following illustrates exemplary possible data. The two haplotypes on same region on the q arm on chromosome 22 are present in a ratio of 4:5; the two haplotypes in a focal region corresponding to the HER2

# US 11,530,454 B2

87

gene are present in ratios of 1:2; and the two haplotypes in the p-arm of chromosome 5 are present in ratios of 20:21. All other assayed regions of the genome have no statistically significant excess of either haplotype. A clinician may infer that the sample contains DNA from a tumor with a CNV in 5 the 22q region, the HER2 region, and the 5p arm. Based on the knowledge that 22q deletions are very common in breast cancer, and/or the quantitative analysis showing an underrepresentation of the amount of DNA mapping to the 22q region of the genome, the clinician may infer the existence 10 of a tumor with a 22q deletion. Based on the knowledge that HER2 amplifications are very common in breast cancer, and/or the quantitative analysis showing an over-representation of the amount of DNA mapping to the HER2 region of the genome, the clinician may infer the existence of a 15 tumor with a HER2 amplification.

Exemplary Reference Chromosomes or Chromosome Segments

In some embodiments, any of the methods described herein are also performed on one or more reference chromosomes or chromosomes segments and the results are compared to those for one or more chromosomes or chromosome segments of interest.

In some embodiments, the reference chromosome or chromosome segment is used as a control for what would be 25 expected for the absence of a CNV. In some embodiments, the reference is the same chromosome or chromosome segment from one or more different samples known or expected to not have a deletion or duplication in that chromosome or chromosome segment. In some embodi- 30 ments, the reference is a different chromosome or chromosome segment from the sample being tested that is expected to be disomic. In some embodiments, the reference is a different segment from one of the chromosomes of interest in the same sample that is being tested. For example, the 3: reference may be one or more segments outside of the region of a potential deletion or duplication. Having a reference on the same chromosome that is being tested avoids variability between different chromosomes, such as differences in metabolism, apoptosis, histones, inactivation, and/or ampli- 40 fication between chromosomes. Analyzing segments without a CNV on the same chromosome as the one being tested can also be used to determine differences in metabolism, apoptosis, hi stones, inactivation, and/or amplification between homologs, allowing the level of variability between 45 homologs in the absence of a CNV to be determined for comparison to the results from a potential CNV. In some embodiments, the magnitude of the difference between the calculated and expected allele ratios for a potential CNV is greater than the corresponding magnitude for the reference, 50 thereby confirming the presence of a CNV.

In some embodiments, the reference chromosome or chromosome segment is used as a control for what would be expected for the presence of a CNV, such as a particular deletion or duplication of interest. In some embodiments, the 55 reference is the same chromosome or chromosome segment from one or more different samples known or expected to have a deletion or duplication in that chromosome or chromosome segment. In some embodiments, the reference is a different chromosome or chromosome segment from the 60 sample being tested that is known or expected to have a CNV. In some embodiments, the magnitude of the difference between the calculated and expected allele ratios for a potential CNV is similar to (such as not significantly different) than the corresponding magnitude for the reference 65 for the CNV, thereby confirming the presence of a CNV. In some embodiments, the magnitude of the difference between

88

the calculated and expected allele ratios for a potential CNV is less than (such as significantly less) than the corresponding magnitude for the reference for the CNV, thereby confirming the absence of a CNV. In some embodiments, one or more loci for which the genotype of a cancer cell (or DNA or RNA from a cancer cell such as cfDNA or cfRNA) differs from the genotype of a noncancerous cell (or DNA or RNA from a noncancerous cell such as cfDNA or cfRNA) is used to determine the tumor fraction. The tumor fraction can be used to determine whether the overrepresentation of the number of copies of the first homologous chromosome segment is due to a duplication of the first homologous chromosome segment or a deletion of the second homologous chromosome segment. The tumor fraction can also be used to determine the number of extra copies of a chromosome segment or chromosome that is duplicated (such as whether there are 1, 2, 3, 4, or more extra copies), such as to differentiate a sample with four extra chromosome copies and a tumor fraction of 10% from a sample with two extra chromosome copies and a tumor fraction of 20%. The tumor fraction can also be used to determine how well the observed data fits the expected data for possible CNVs. In some embodiments, the degree of overrepresentation of a CNV is used to select a particular therapy or therapeutic regimen for the individual. For example, some therapeutic agents are only effective for at least four, six, or more copies of a chromosome segment.

In some embodiments, the one or more loci used to determine the tumor fraction are on a reference chromosome or chromosomes segment, such as a chromosome or chromosome segment known or expected to be disomic, a chromosome or chromosome segment that is rarely duplicated or deleted in cancer cells in general or in a particular type of cancer that an individual is known to have or is at increased risk of having, or a chromosome or chromosome segment that is unlikely to be an uploid (such segment that is expected to lead to cell death if deleted or duplicated). In some embodiments, any of the methods of the invention are used to confirm that the reference chromosome or chromosome segment is disomic in both the cancer cells and noncancerous cells. In some embodiments, one or more chromosomes or chromosomes segments for which the confidence for a disomy call is high are used.

Exemplary loci that can be used to determine the tumor fraction include polymorphisms or mutations (such as SNPs) in a cancer cell (or DNA or RNA such as cfDNA or cfRNA from a cancer cell) that aren't present in a noncancerous cell (or DNA or RNA from a noncancerous cell) in the individual. In some embodiments, the tumor fraction is determined by identifying those polymorphic loci where a cancer cell (or DNA or RNA from a cancer cell) has an allele that is absent in noncancerous cells (or DNA or RNA from a noncancerous cell) in a sample (such as a plasma sample or tumor biopsy) from an individual; and using the amount of the allele unique to the cancer cell at one or more of the identified polymorphic loci to determine the tumor fraction in the sample. In some embodiments, a noncancerous cell is homozygous for a first allele at the polymorphic locus, and a cancer cell is (i) heterozygous for the first allele and a second allele or (ii) homozygous for a second allele at the polymorphic locus. In some embodiments, a noncancerous cell is heterozygous for a first allele and a second allele at the polymorphic locus, and a cancer cell is (i) has one or two copies of a third allele at the polymorphic locus. In some embodiments, the cancer cells are assumed or known to only have one copy of the allele that is not present in the noncancerous cells. For example, if the genotype of the

# US 11,530,454 B2

89

noncancerous cells is AA and the cancer cells is AB and 5% of the signal at that locus in a sample is from the B allele and 95% is from the A allele, then the tumor fraction of the sample is 10%. In some embodiments, the cancer cells are assumed or known to have two copies of the allele that is not 5 present in the noncancerous cells. For example, if the genotype of the noncancerous cells is AA and the cancer cells is BB and 5% of the signal at that locus in a sample is from the B allele and 95% is from the A allele, the tumor fraction of the sample is 5%. In some embodiments, multiple 10 loci for which the cancer cells have an allele not in the noncancerous cells are analyzed to determine which of the loci in the cancer cells are heterozygous and which are homozygous. For example for loci in which the noncancerous cells are AA, if the signal from the B allele is ~5% at 15 some loci and ~10% at some loci, then the cancer cells are assumed to be heterozygous at loci with ~5% B allele, and homozygous at loci with ~10% B allele (indicating the tumor fraction is ~10%).

Exemplary loci that can be used to determine the tumor 20 fraction include loci for which a cancer cell and noncancerous cell have one allele in common (such as loci in which the cancer cell is AB and the noncancerous cell is BB, or the cancer cell is BB and the noncancerous cell is AB). The amount of A signal, the amount of B signal, or the ratio of 25 A to B signal in a mixed sample (containing DNA or RNA from a cancer cell and a noncancerous cell) is compared to the corresponding value for (i) a sample containing DNA or RNA from only cancer cells or (ii) a sample containing DNA or RNA from only noncancerous cells. The difference in 30 values is used to determine the tumor fraction of the mixed sample.

In some embodiments, loci that can be used to determine the tumor fraction are selected based on the genotype of (i) a sample containing DNA or RNA from only cancer cells, 3: and/or (ii) a sample containing DNA or RNA from only noncancerous cells. In some embodiments, the loci are selected based on analysis of the mixed sample, such as loci for which the absolute or relative amounts of each allele differs from what would be expected if both the cancer and 40 noncancerous cells have the same genotype at a particular locus. For example, if the cancer and noncancerous cells have the same genotype, the loci would be expected to produce 0% B signal if all the cells are AA, 50% B signal if all the cells are AB, or 100% B signal if all the cells are 45 BB. Other values for the B signal indicate that the genotype of the cancer and noncancerous cells are different at that locus and thus that locus can be used to determine the tumor fraction.

In some embodiments, the tumor fraction calculated 50 based on the alleles at one or more loci is compared to the tumor fraction calculated using one or more of the counting methods disclosed herein.

Exemplary Methods for Detecting a Phenotype or Analyzing Multiple Mutations

In some embodiments, the method includes analyzing a sample for a set of mutations associated with a disease or disorder (such as cancer) or an increased risk for a disease or disorder. There are strong correlations between events within classes (such as M or C cancer classes) which can be used to improve the signal to noise ratio of a method and classify tumors into distinct clinical subsets. For example, borderline results for a few mutations (such as a few CNVs) on one or more chromosomes or chromosomes segments considered jointly may be a very strong signal. In some 65 embodiments, determining the presence or absence of multiple polymorphisms or mutations of interest (such as 2, 3,

90

4, 5, 8, 10, 12, 15, or more) increases the sensitivity and/or specificity of the determination of the presence or absence of a disease or disorder such as cancer, or an increased risk for with a disease or disorder such as cancer. In some embodiments, the correlation between events across multiple chromosomes is used to more powerfully look at a signal compared to looking at each of them individually. The design of the method itself can be optimized to best categorize tumors. This may be incredibly useful for early detection and screening—vis-a-vis recurrence where sensitivity to one particular mutation/CNV may be paramount. In some embodiments, the events are not always correlated but have a probability of being correlated. In some embodiments, a matrix estimation formulation with a noise covariance matrix that has off diagonal terms is used.

In some embodiments, the invention features a method for detecting a phenotype (such as a cancer phenotype) in an individual, wherein the phenotype is defined by the presence of at least one of a set of mutations. In some embodiments. the method includes obtaining DNA or RNA measurements for a sample of DNA or RNA from one or more cells from the individual, wherein one or more of the cells is suspected of having the phenotype; and analyzing the DNA or RNA measurements to determine, for each of the mutations in the set of mutations, the likelihood that at least one of the cells has that mutation. In some embodiments, the method includes determining that the individual has the phenotype if either (i) for at least one of the mutations, the likelihood that at least one of the cells contains that mutations is greater than a threshold, or (ii) for at least one of the mutations, the likelihood that at least one of the cells has that mutations is less than the threshold, and for a plurality of the mutations, the combined likelihood that at least one of the cells has at least one of the mutations is greater than the threshold. In some embodiments, one or more cells have a subset or all of the mutations in the set of mutations. In some embodiments, the subset of mutations is associated with cancer or an increased risk for cancer. In some embodiments, the set of mutations includes a subset or all of the mutations in the M class of cancer mutations (Ciriello, Nat Genet. 45(10):1127-1133, 2013, doi: 10.1038/ng.2762, which is hereby incorporated by reference in its entirety). In some embodiments, the set of mutations includes a subset or all of the mutations in the C class of cancer mutations (Ciriello, supra). In some embodiments, the sample includes cell-free DNA or RNA. In some embodiments, the DNA or RNA measurements include measurements (such as the quantity of each allele at each locus) at a set of polymorphic loci on one or more chromosomes or chromosome segments of interest.

Exemplary Methods for Paternity Testing or Genetic Relatedness Testing

The methods of the invention can be used to improve the accuracy of paternity testing or other genetic relatedness testing (see, e.g., U.S. Publication No. 2012/0122701, filed Dec. 22, 2011, which is hereby incorporated by reference in its entirety). For example, the multiplex PCR method can allow thousands of polymorphic loci (such as SNPs) to be analyzed for use in the PARENTAL SUPPORT algorithm described herein to determine whether an alleged father in is the biological father of a fetus. In some embodiments, the invention features a method for establishing whether an alleged father is the biological father of a fetus that is gestating in a pregnant mother. In some embodiments, the method involves obtaining phased genetic data for the alleged father (such as by using another of the methods described herein for phasing genetic data), wherein the phased genetic data comprises the identity of the allele

# US 11,530,454 B2

9

present for each locus in a set of polymorphic loci on a first homologous chromosome segment and a second homologous chromosome segment in the alleged father. In some embodiments, the method involves obtaining genetic data at the set of polymorphic loci on the chromosome or chromosome segment in a mixed sample of DNA comprising fetal DNA and maternal DNA from the mother of the fetus by measuring the quantity of each allele at each locus. In some embodiments, the method involves calculating, on a computer, expected genetic data for the mixed sample of DNA 10 from the phased genetic data for the alleged father; determining, on a computer, the probability that the alleged father is the biological father of the fetus by comparing the obtaining genetic data made on the mixed sample of DNA to the expected genetic data for the mixed sample of DNA; 15 and establishing whether the alleged father is the biological father of the fetus using the determined probability that the alleged father is the biological father of the fetus. In some embodiments, the method involves obtaining phased genetic data for the biological mother of the fetus (such as by using 20 another of the methods described herein for phasing genetic data), wherein the phased genetic data comprises the identity of the allele present for each locus in a set of polymorphic loci on a first homologous chromosome segment and a second homologous chromosome segment in the mother. In 25 some embodiments, the method involves obtaining phased genetic data for the fetus (such as by using another of the methods described herein for phasing genetic data), wherein the phased genetic data comprises the identity of the allele present for each locus in a set of polymorphic loci on a first 30 homologous chromosome segment and a second homologous chromosome segment in the fetus. In some embodiments, the method involves calculating, on a computer, expected genetic data for the mixed sample of DNA using the phased genetic data for the alleged father and using the 35 phased genetic data for the mother and/or the phased genetic data for the fetus.

In some embodiments, the invention features a method for establishing whether an alleged father is the biological father of a fetus that is gestating in a pregnant mother. In some 40 embodiments, the method involves obtaining phased genetic data for the alleged father (such as by using another of the methods described herein for phasing genetic data), wherein the phased genetic data comprises the identity of the allele present for each locus in a set of polymorphic loci on a first 45 homologous chromosome segment and a second homologous chromosome segment in the alleged father. In some embodiments, the method involves obtaining genetic data at the set of polymorphic loci on the chromosome or chromosome segment in a mixed sample of DNA comprising fetal 50 DNA and maternal DNA from the mother of the fetus by measuring the quantity of each allele at each locus. In some embodiments, the method involves identifying (i) alleles that are present in the fetal DNA but are absent in the maternal DNA at polymorphic loci, and/or identifying (i) 55 alleles that are absent in the fetal DNA and the maternal DNA at polymorphic loci. In some embodiments, the method involves determining, on a computer, the probability that the alleged father is the biological father of the fetus; wherein the determination comprises: (1) comparing (i) the 60 alleles that are present in the fetal DNA but are absent in the maternal DNA at polymorphic loci to (ii) the alleles at the corresponding polymorphic loci in the genetic material from the alleged father, and/or (2) comparing (i) the alleles that are absent in the fetal DNA and the maternal DNA at 65 polymorphic loci to (ii) the alleles at the corresponding polymorphic loci in the genetic material from the alleged

92

father; and establishing whether the alleged father is the biological father of the fetus using the determined probability that the alleged father is the biological father of the fetus.

In some embodiments, a method described above for determining whether an alleged father is the biological father of the fetus is used to determine if an alleged relative (such as a grandparent, sibling, aunt, or uncle) of a fetus is an actual biological relative of the fetus (such as by using genetic data of the alleged relative instead of genetic data of the alleged father).

**Exemplary Combinations of Methods** 

To increase the accuracy of the results, two or more methods (such as any of the methods of the invention or any known method) for detecting the presence or absence of a CNV are performed. In some embodiments, one or more methods for analyzing a factor (such as any of the method described herein or any known method) indicative of the presence or absence of a disease or disorder or an increased risk for a disease or disorder are performed.

In some embodiments, standard mathematical techniques are used to calculate the covariance and/or correlation between two or more methods. Standard mathematical techniques may also be used to determine the combined probability of a particular hypothesis based on two or more tests. Exemplary techniques include meta-analysis, Fisher's combined probability test for independent tests, Brown's method for combining dependent p-values with known covariance, and Kost's method for combining dependent p-values with unknown covariance. In cases where the likelihoods are determined by a first method in a way that is orthogonal, or unrelated, to the way in which a likelihood is determined for a second method, combining the likelihoods is straightforward and can be done by multiplication and normalization, or by using a formula such as:

$$R_{comb} = R_1 R_2 / [R_1 R_2 + (1 - R_1)(1 - R_2)]$$

 $R_{comb}$  is the combined likelihood, and  $R_1$  and  $R_2$  are the individual likelihoods. For example, if the likelihood of trisomy from method 1 is 90%, and the likelihood of trisomy from method 2 is 95%, then combining the outputs from the two methods allows the clinician to conclude that the fetus is trisomic with a likelihood of (0.90)(0.95)/[(0.90)(0.95)+(1-0.90)(1-0.95)]=99.42%. In cases where the first and the second methods are not orthogonal, that is, where there is a correlation between the two methods, the likelihoods can still be combined.

Exemplary methods of analyzing multiple factors or variables are disclosed in U.S. Pat. No. 8,024,128 issued on Sep. 20, 2011; U.S. Publication No. 2007/0027636, filed Jul. 31, 2006; and U.S. Publication No. 2007/0178501, filed Dec. 6, 2006, which are each hereby incorporated by reference in its entirety).

In various embodiments, the combined probability of a particular hypothesis or diagnosis is greater than 80, 85, 90, 92, 94, 96, 98, 99, or 99.9%, or is greater than some other threshold value.

Limit of Detection

In some embodiments, a limit of detection of a mutation (such as an SNV or CNV) of a method of the invention is less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005%. In some embodiments, a limit of detection of a mutation (such as an SNV or CNV) of a method of the invention is between 15 to 0.005%, such as between 10 to 0.005%, 10 to 0.01%, 10 to 0.1%, 5 to 0.005%, 5 to 0.01%, 5 to 0.1%, 1 to 0.005%, 1 to 0.005%, 0.5 to 0.01%, 0.5 to 0.01%, 0.5 to 0.01%, and 0.5 to 0.01%, 0.5 to 0.01%, and 0.5 to 0.01%, or 0.1 to 0.01, inclusive. In some embodiments, a limit of detection is such that a

Case: 24-1324 Document: 42-1 Page: 496 Filed: 03/18/2024

# US 11,530,454 B2

mutation (such as an SNV or CNV) that is present in less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules with that locus in a sample (such as a sample of cfDNA or cfRNA) is detected (or is capable of being detected). For example, the mutation can be detected even if less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules that have that locus have that mutation in the locus (instead of, for example, a wild-type or non-mutated version of the locus or a different mutation at that locus). In some embodiments, 10 a limit of detection is such that a mutation (such as an SNV or CNV) that is present in less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample (such as a sample of cfDNA or cfRNA) is detected (or is capable of being detected). In 15 some embodiments in which the CNV is a deletion, the deletion can be detected even if it is only present in less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules that have a region of interest that may or may not contain the deletion in a sample. In some 20 embodiments in which the CNV is a deletion, the deletion can be detected even if it is only present in less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample. In some embodiments in which the CNV is a duplication, the duplication can be detected 25 cells. In some embodiments in which the sample is enriched even if the extra duplicated DNA or RNA that is present is less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules that have a region of interest that may or may not be duplicated in a sample in a sample. In some embodiments in which the CNV is a 30 duplication, the duplication can be detected even if the extra duplicated DNA or RNA that is present is less than or equal to 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample. Example 6 provides exemplary methods for calculating the limit of detection. In some 35 embodiments, the "LOD-zs5.0-mr5" method of Example 6 is used.

### **Exemplary Samples**

In some embodiments of any of the aspects of the invention, the sample includes cellular and/or extracellular 40 genetic material from cells suspected of having a deletion or duplication, such as cells suspected of being cancerous. In some embodiments, the sample comprises any tissue or bodily fluid suspected of containing cells, DNA, or RNA having a deletion or duplication, such as cancer cells, DNA, 45 or RNA. The genetic measurements used as part of these methods can be made on any sample comprising DNA or RNA, for example but not limited to, tissue, blood, serum, plasma, urine, hair, tears, saliva, skin, fingernails, feces, bile, lymph, cervical mucus, semen, or other cells or materials 50 comprising nucleic acids. Samples may include any cell type or DNA or RNA from any cell type may be used (such as cells from any organ or tissue suspected of being cancerous, or neurons). In some embodiments, the sample includes nuclear and/or mitochondrial DNA. In some embodiments, 55 the sample is from any of the target individuals disclosed herein. In some embodiments, the target individual is a born individual, a gestating fetus, a non-gestating fetus such as a products of conception sample, an embryo, or any other

Exemplary samples include those containing cfDNA or cfRNA. In some embodiments, cfDNA is available for analysis without requiring the step of lysing cells. Cell-free DNA may be obtained from a variety of tissues, such as tissues that are in liquid form, e.g., blood, plasma, lymph, 65 ascites fluid, or cerebral spinal fluid. In some cases, cfDNA is comprised of DNA derived from fetal cells. In some cases,

cfDNA is comprised of DNA derived from both fetal and maternal cells. In some cases, the cfDNA is isolated from plasma that has been isolated from whole blood that has been centrifuged to remove cellular material. The cfDNA may be a mixture of DNA derived from target cells (such as cancer cells) and non-target cells (such as non-cancer cells).

In some embodiments, the sample contains or is suspected to contain a mixture of DNA (or RNA), such as mixture of cancer DNA (or RNA) and noncancerous DNA (or RNA). In some embodiments, at least 0.5, 1, 3, 5, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 92, 94, 95, 96, 98, 99, or 100% of the cells in the sample are cancer cells. In some embodiments, at least 0.5, 1, 3, 5, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 92, 94, 95, 96, 98, 99, or 100% of the DNA (such as cfDNA) or RNA (such as cfRNA) in the sample is from cancer cell(s). In various embodiments, the percent of cells in the sample that are cancerous cells is between 0.5 to 99%, such as between 1 to 95%, 5 to 95%, 10 to 90%, 5 to 70%, 10 to 70%, 20 to 90%, or 20 to 70%, inclusive. In some embodiments, the sample is enriched for cancer cells or for DNA or RNA from cancer cells. In some embodiments in which the sample is enriched for cancer cells, at least 0.5, 1, 2, 3, 4, 5, 6, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 92, 94, 95, 96, 98, 99, or 100% of the cells in the enriched sample are cancer for DNA or RNA from cancer cells, at least 0.5, 1, 2, 3, 4, 5, 6, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 92, 94, 95, 96, 98, 99, or 100% of the DNA or RNA in the enriched sample is from cancer cell(s). In some embodiments, cell sorting (such as Fluorescent Activated Cell Sorting (FACS)) is used to enrich for cancer cells (Barteneva et. al., Biochim Biophys Acta., 1836(1):105-22, August 2013. doi: 10.1016/ j.bbcan.2013.02.004. Epub 2013 Feb. 24, and Ibrahim et al., Adv Biochem Eng Biotechnol. 106:19-39, 2007, which are each hereby incorporated by reference in its entirety).

In some embodiments of any of the aspects of the invention, the sample comprises any tissue suspected of being at least partially of fetal origin. In some embodiments, the sample includes cellular and/or extracellular genetic material from the fetus, contaminating cellular and/or extracellular genetic material (such as genetic material from the mother of the fetus), or combinations thereof. In some embodiments, the sample comprises cellular genetic material from the fetus, contaminating cellular genetic material, or combinations thereof.

In some embodiments, the sample is from a gestating fetus. In some embodiments, the sample is from a nongestating fetus, such as a products of conception sample or a sample from any fetal tissue after fetal demise. In some embodiments, the sample is a maternal whole blood sample, cells isolated from a maternal blood sample, maternal plasma sample, maternal serum sample, amniocentesis sample, placental tissue sample (e.g., chorionic villus, decidua, or placental membrane), cervical mucus sample, or other sample from a fetus. In some embodiments, at least 3, 5, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 92, 94, 95, 96, 98, 99, or 100% of the cells in the sample are maternal cells. In various embodiments, the percent of cells in the sample that are maternal cells is between 5 to 99%, such as between 10 to 95%, 20 to 95%, 30 to 90%, 30 to 70%, 40 to 90%, 40 to 70%, 50 to 90%, or 50 to 80%, inclusive.

In some embodiments, the sample is enriched for fetal cells. In some embodiments in which the sample is enriched for fetal cells, at least 0.5, 1, 2, 3, 4, 5, 6, 7% or more of the cells in the enriched sample are fetal cells. In some embodiments, the percent of cells in the sample that are fetal cells is between 0.5 to 100%, such as between 1 to 99%, 5 to 95%,

# US 11,530,454 B2

9:

10 to 95%, 10 to 95%, 20 to 90%, or 30 to 70%, inclusive. In some embodiments, the sample is enriched for fetal DNA. In some embodiments in which the sample is enriched for fetal DNA, at least 0.5, 1, 2, 3, 4, 5, 6, 7% or more of the DNA in the enriched sample is fetal DNA. In some embodiments, the percent of DNA in the sample that is fetal DNA is between 0.5 to 100%, such as between 1 to 99%, 5 to 95%, 10 to 95%, 10 to 95%, 20 to 90%, or 30 to 70%, inclusive.

In some embodiments, the sample includes a single cell or includes DNA and/or RNA from a single cell. In some 10 embodiments, multiple individual cells (e.g., at least 5, 10, 20, 30, 40, or 50 cells from the same subject or from different subjects) are analyzed in parallel. In some embodiments, cells from multiple samples from the same individual are combined, which reduces the amount of work compared to analyzing the samples separately. Combining multiple samples can also allow multiple tissues to be tested for cancer simultaneously (which can be used to provide or more thorough screening for cancer or to determine whether cancer may have metastasized to other tissues).

In some embodiments, the sample contains a single cell or a small number of cells, such as 2, 3, 5, 6, 7, 8, 9, or 10 cells. In some embodiments, the sample has between 1 to 100, 100 to 500, or 500 to 1,000 cells, inclusive. In some embodiments, the sample contains 1 to 10 picograms, 10 to 100 picograms to 1 nanogram, 1 to 10 nanograms, 10 to 100 nanograms, or 100 nanograms to 1 microgram of RNA and/or DNA, inclusive.

In some embodiments, the sample is embedded in parafilm. In some embodiments, the sample is preserved with a preservative such as formaldehyde and optionally encased in paraffin, which may cause cross-linking of the DNA such that less of it is available for PCR. In some embodiments, the sample is a formaldehyde fixed-paraffin embedded (FFPE) sample. In some embodiments, the sample is a fresh sample (such as a sample obtained with 1 or 2 days of analysis). In some embodiments, the sample is frozen prior to analysis. In some embodiments, the sample is a historical sample.

These samples can be used in any of the methods of the invention.

Exemplary Sample Preparation Methods

In some embodiments, the method includes isolating or purifying the DNA and/or RNA. There are a number of standard procedures known in the art to accomplish such an end. In some embodiments, the sample may be centrifuged 45 to separate various layers. In some embodiments, the DNA or RNA may be isolated using filtration. In some embodiments, the preparation of the DNA or RNA may involve amplification, separation, purification by chromatography, liquid liquid separation, isolation, preferential enrichment, 50 preferential amplification, targeted amplification, or any of a number of other techniques either known in the art or described herein. In some embodiments for the isolation of DNA, RNase is used to degrade RNA. In some embodiments for the isolation of RNA, DNase (such as DNase I 55 from Invitrogen, Carlsbad, Calif., USA) is used to degrade DNA. In some embodiments, an RNeasy mini kit (Qiagen), is used to isolate RNA according to the manufacturer's protocol. In some embodiments, small RNA molecules are isolated using the mirVana PARIS kit (Ambion, Austin, Tex., 60 USA) according to the manufacturer's protocol (Gu et al., J. Neurochem. 122:641-649, 2012, which is hereby incorporated by reference in its entirety). The concentration and purity of RNA may optionally be determined using Nanovue (GE Healthcare, Piscataway, N.J., USA), and RNA integrity 65 may optionally be measured by use of the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, Calif., USA) (Gu et al.,

96

J. Neurochem. 122:641-649, 2012, which is hereby incorporated by reference in its entirety). In some embodiments, TRIZOL or RNAlater (Ambion) is used to stabilize RNA during storage.

In some embodiments, universal tagged adaptors are added to make a library. Prior to ligation, sample DNA may be blunt ended, and then a single adenosine base is added to the 3-prime end. Prior to ligation the DNA may be cleaved using a restriction enzyme or some other cleavage method. During ligation the 3-prime adenosine of the sample fragments and the complementary 3-prime tyrosine overhang of adaptor can enhance ligation efficiency. In some embodiments, adaptor ligation is performed using the ligation kit found in the AGILENT SURESELECT kit. In some embodiments, the library is amplified using universal primers. In an embodiment, the amplified library is fractionated by size separation or by using products such as AGENCOURT AMPURE beads or other similar methods. In some embodiments, PCR amplification is used to amplify target loci. In 20 some embodiments, the amplified DNA is sequenced (such as sequencing using an ILLUMINA IIGAX or HiSeq sequencer). In some embodiments, the amplified DNA is sequenced from each end of the amplified DNA to reduce sequencing errors. If there is a sequence error in a particular base when sequencing from one end of the amplified DNA, there is less likely to be a sequence error in the complementary base when sequencing from the other side of the amplified DNA (compared to sequencing multiple times from the same end of the amplified DNA).

In some embodiments, whole genome application (WGA) is used to amplify a nucleic acid sample. There are a number of methods available for WGA: ligation-mediated PCR (LM-PCR), degenerate oligonucleotide primer PCR (DOP-PCR), and multiple displacement amplification (MDA). In LM-PCR, short DNA sequences called adapters are ligated to blunt ends of DNA. These adapters contain universal amplification sequences, which are used to amplify the DNA by PCR. In DOP-PCR, random primers that also contain universal amplification sequences are used in a first round of 40 annealing and PCR. Then, a second round of PCR is used to amplify the sequences further with the universal primer sequences. MDA uses the phi-29 polymerase, which is a highly processive and non-specific enzyme that replicates DNA and has been used for single-cell analysis. In some embodiments, WGA is not performed.

In some embodiments, selective amplification or enrichment are used to amplify or enrich target loci. In some embodiments, the amplification and/or selective enrichment technique may involve PCR such as ligation mediated PCR, fragment capture by hybridization, Molecular Inversion Probes, or other circularizing probes. In some embodiments, real-time quantitative PCR (RT-qPCR), digital PCR, or emulsion PCR, single allele base extension reaction followed by mass spectrometry are used (Hung et al., J Clin Pathol 62:308-313, 2009, which is hereby incorporated by reference in its entirety). In some embodiments, capture by hybridization with hybrid capture probes is used to preferentially enrich the DNA. In some embodiments, methods for amplification or selective enrichment may involve using probes where, upon correct hybridization to the target sequence, the 3-prime end or 5-prime end of a nucleotide probe is separated from the polymorphic site of a polymorphic allele by a small number of nucleotides. This separation reduces preferential amplification of one allele, termed allele bias. This is an improvement over methods that involve using probes where the 3-prime end or 5-prime end of a correctly hybridized probe are directly adjacent to or very

# US 11,530,454 B2

97

near to the polymorphic site of an allele. In an embodiment, probes in which the hybridizing region may or certainly contains a polymorphic site are excluded. Polymorphic sites at the site of hybridization can cause unequal hybridization or inhibit hybridization altogether in some alleles, resulting 5 in preferential amplification of certain alleles. These embodiments are improvements over other methods that involve targeted amplification and/or selective enrichment in that they better preserve the original allele frequencies of the sample at each polymorphic locus, whether the sample 10 is pure genomic sample from a single individual or mixture of individuals

In some embodiments, PCR (referred to as mini-PCR) is used to generate very short amplicons (U.S. application Ser. No. 13/683,604, filed Nov. 21, 2012, U.S. Publication No. 15 2013/0123120, U.S. application Ser. No. 13/300,235, filed Nov. 18, 2011, U.S. Publication No 2012/0270212, filed Nov. 18, 2011, and U.S. Ser. No. 61/994,791, filed May 16, 2014, which are each hereby incorporated by reference in its entirety). cfDNA (such as fetal cfDNA in maternal serum or 20 necroptically- or apoptotically-released cancer cfDNA) is highly fragmented. For fetal cfDNA, the fragment sizes are distributed in approximately a Gaussian fashion with a mean of 160 bp, a standard deviation of 15 bp, a minimum size of about 100 bp, and a maximum size of about 220 bp. The 25 polymorphic site of one particular target locus may occupy any position from the start to the end among the various fragments originating from that locus. Because cfDNA fragments are short, the likelihood of both primer sites being present the likelihood of a fragment of length L comprising both the forward and reverse primers sites is the ratio of the length of the amplicon to the length of the fragment. Under ideal conditions, assays in which the amplicon is 45, 50, 55, 60, 65, or 70 bp will successfully amplify from 72%, 69%, 66%, 63%, 59%, or 56%, respectively, of available template 3. fragment molecules. In certain embodiments that relate most preferably to cfDNA from samples of individuals suspected of having cancer, the cfDNA is amplified using primers that yield a maximum amplicon length of 85, 80, 75 or 70 bp, and in certain preferred embodiments 75 bp, and that have a 4 melting temperature between 50 and 65° C., and in certain preferred embodiments, between 54-60.5° C. The amplicon length is the distance between the 5-prime ends of the forward and reverse priming sites. Amplicon length that is shorter than typically used by those known in the art may result in more efficient measurements of the desired polymorphic loci by only requiring short sequence reads. In an embodiment, a substantial fraction of the amplicons are less than 100 bp, less than 90 bp, less than 80 bp, less than 70 bp, less than 65 bp, less than 60 bp, less than 55 bp, less than 50 50 bp, or less than 45 bp.

In some embodiments, amplification is performed using direct multiplexed PCR, sequential PCR, nested PCR, doubly nested PCR, one-and-a-half sided nested PCR, fully nested PCR, one sided fully nested PCR, one-sided nested 55 PCR, hemi-nested PCR, hemi-nested PCR, triply hemi-nested PCR, semi-nested PCR, one sided semi-nested PCR, reverse semi-nested PCR method, or one-sided PCR, which are described in U.S. application Ser. No. 13/683,604, filed Nov. 21, 2012, U.S. Publication No. 2013/0123120, U.S. application Ser. No. 13/300,235, filed Nov. 18, 2011, U.S. Publication No 2012/0270212, and U.S. Ser. No. 61/994, 791, filed May 16, 2014, which are hereby incorporated by reference in their entirety. If desired, any of these methods can be used for mini-PCR.

If desired, the extension step of the PCR amplification may be limited from a time standpoint to reduce amplifica98

tion from fragments longer than 200 nucleotides, 300 nucleotides, 400 nucleotides, 500 nucleotides or 1,000 nucleotides. This may result in the enrichment of fragmented or shorter DNA (such as fetal DNA or DNA from cancer cells that have undergone apoptosis or necrosis) and improvement of test performance.

In some embodiments, multiplex PCR is used. In some embodiments, the method of amplifying target loci in a nucleic acid sample involves (i) contacting the nucleic acid sample with a library of primers that simultaneously hybridize to least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci to produce a reaction mixture; and (ii) subjecting the reaction mixture to primer extension reaction conditions (such as PCR conditions) to produce amplified products that include target amplicons. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the targeted loci are amplified. In various embodiments, less than 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.05% of the amplified products are primer dimers. In some embodiments, the primers are in solution (such as being dissolved in the liquid phase rather than in a solid phase). In some embodiments, the primers are in solution and are not immobilized on a solid support. In some embodiments, the primers are not part of a microarray. In some embodiments, the primers do not include molecular inversion probes (MIPs)

In some embodiments, two or more (such as 3 or 4) target amplicons (such as amplicons from the miniPCR method disclosed herein) are ligated together and then the ligated products are sequenced. Combining multiple amplicons into a single ligation product increases the efficiency of the subsequent sequencing step. In some embodiments, the target amplicons are less than 150, 100, 90, 75, or 50 base pairs in length before they are ligated. The selective enrichment and/or amplification may involve tagging each individual molecule with different tags, molecular barcodes, tags for amplification, and/or tags for sequencing. In some embodiments, the amplified products are analyzed by sequencing (such as by high throughput sequencing) or by hybridization to an array, such as a SNP array, the ILLU-MINA INFINIUM array, or the AFFYMETRIX gene chip. In some embodiments, nanopore sequencing is used, such as the nanopore sequencing technology developed by Genia (see, for example, the world wide web at geniachip.com/ technology, which is hereby incorporated by reference in its entirety). In some embodiments, duplex sequencing is used (Schmitt et al., "Detection of ultra-rare mutations by nextgeneration sequencing," Proc Natl Acad Sci USA. 109(36): 14508-14513, 2012, which is hereby incorporated by reference in its entirety). This approach greatly reduces errors by independently tagging and sequencing each of the two strands of a DNA duplex. As the two strands are complementary, true mutations are found at the same position in both strands. In contrast, PCR or sequencing errors result in mutations in only one strand and can thus be discounted as technical error. In some embodiments, the method entails tagging both strands of duplex DNA with a random, yet complementary double-stranded nucleotide sequence, referred to as a Duplex Tag. Double-stranded tag sequences are incorporated into standard sequencing adapters by first introducing a single-stranded randomized nucleotide sequence into one adapter strand and then extending the opposite strand with a DNA polymerase to yield a complementary, double-stranded tag. Following ligation of tagged adapters to sheared DNA, the individually labeled strands are PCR amplified from asymmetric primer sites on the Case: 24-1324 Document: 42-1 Page: 499 Filed: 03/18/2024

# US 11,530,454 B2

adapter tails and subjected to paired-end sequencing. In some embodiments, a sample (such as a DNA or RNA sample) is divided into multiple fractions, such as different wells (e.g., wells of a WaferGen SmartChip). Dividing the sample into different fractions (such as at least 5, 10, 20, 50, 75, 100, 150, 200, or 300 fractions) can increase the sensitivity of the analysis since the percent of molecules with a mutation are higher in some of the wells than in the overall sample. In some embodiments, each fraction has less than 500, 400, 200, 100, 50, 20, 10, 5, 2, or 1 DNA or RNA 10 molecules. In some embodiments, the molecules in each fraction are sequenced separately. In some embodiments, the same barcode (such as a random or non-human sequence) is added to all the molecules in the same fraction (such as by amplification with a primer containing the barcode or by 15 ligation of a barcode), and different barcodes are added to molecules in different fractions. The barcoded molecules can be pooled and sequenced together. In some embodiments, the molecules are amplified before they are pooled and ments, one forward and two reverse primers, or two forward and one reverse primers are used.

In some embodiments, a mutation (such as an SNV or CNV) that is present in less than 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, or 0.005% of the DNA or RNA molecules in a sample 25 (such as a sample of cfDNA or cfRNA) is detected (or is capable of being detected). In some embodiments, a mutation (such as an SNV or CNV) that is present in less than 1,000, 500, 100, 50, 20, 10, 5, 4, 3, or 2 original DNA or RNA molecules (before amplification) in a sample (such as 30 a sample of cfDNA or cfRNA from, e.g., a blood sample) is detected (or is capable of being detected). In some embodiments, a mutation (such as an SNV or CNV) that is present in only 1 original DNA or RNA molecule (before amplification) in a sample (such as a sample of cfDNA or cfRNA 35 from, e.g., a blood sample) is detected (or is capable of being detected).

For example, if the limit of detection of a mutation (such as a single nucleotide variant (SNV)) is 0.1%, a mutation present at 0.01% can be detected by dividing the fraction 40 into multiple, fractions such as 100 wells. Most of the wells have no copies of the mutation. For the few wells with the mutation, the mutation is at a much higher percentage of the reads. In one example, there are 20,000 initial copies of DNA from the target locus, and two of those copies include 45 a SNV of interest. If the sample is divided into 100 wells, 98 wells have the SNV, and 2 wells have the SNV at 0.5%. The DNA in each well can be barcoded, amplified, pooled with DNA from the other wells, and sequenced. Wells without the SNV can be used to measure the background amplification/ 50 sequencing error rate to determine if the signal from the outlier wells is above the background level of noise.

In some embodiments, the amplified products are detected using an array, such as an array especially a microarray with probes to one or more chromosomes of interest (e.g., chro-55 mosome 13, 18, 21, X, Y, or any combination thereof). It will be understood for example, that a commercially available SNP detection microarray could be used such as, for example, the Illumina (San Diego, Calif.) GoldenGate, DASL, Infinium, or CytoSNP-12 genotyping assay, or a 60 SNP detection microarray product from Affymetrix, such as the OncoScan microarray. In some embodiments, phased genetic data for one or both biological parents of the embryo or fetus is used to increase the accuracy of analysis of array data from a single cell.

In some embodiments involving sequencing, the depth of read is the number of sequencing reads that map to a given

locus. The depth of read may be normalized over the total number of reads. In some embodiments for depth of read of a sample, the depth of read is the average depth of read over the targeted loci. In some embodiments for the depth of read of a locus, the depth of read is the number of reads measured by the sequencer mapping to that locus. In general, the greater the depth of read of a locus, the closer the ratio of alleles at the locus tend to be to the ratio of alleles in the original sample of DNA. Depth of read can be expressed in variety of different ways, including but not limited to the percentage or proportion. Thus, for example in a highly parallel DNA sequencer such as an Illumina HISEQ, which, e.g., produces a sequence of 1 million clones, the sequencing of one locus 3,000 times results in a depth of read of 3,000 reads at that locus. The proportion of reads at that locus is 3,000 divided by 1 million total reads, or 0.3% of the total

In some embodiments, allelic data is obtained, wherein the allelic data includes quantitative measurement(s) indicasequenced, such as by using nested PCR. In some embodi- 20 tive of the number of copies of a specific allele of a polymorphic locus. In some embodiments, the allelic data includes quantitative measurement(s) indicative of the number of copies of each of the alleles observed at a polymorphic locus. Typically, quantitative measurements are obtained for all possible alleles of the polymorphic locus of interest. For example, any of the methods discussed in the preceding paragraphs for determining the allele for a SNP or SNV locus, such as for example, microarrays, qPCR, DNA sequencing, such as high throughput DNA sequencing, can be used to generate quantitative measurements of the number of copies of a specific allele of a polymorphic locus. This quantitative measurement is referred to herein as allelic frequency data or measured genetic allelic data. Methods using allelic data are sometimes referred to as quantitative allelic methods; this is in contrast to quantitative methods which exclusively use quantitative data from non-polymorphic loci, or from polymorphic loci but without regard to allelic identity. When the allelic data is measured using high-throughput sequencing, the allelic data typically include the number of reads of each allele mapping to the

In some embodiments, non-allelic data is obtained, wherein the non-allelic data includes quantitative measurement(s) indicative of the number of copies of a specific locus. The locus may be polymorphic or non-polymorphic. In some embodiments when the locus is non-polymorphic, the non-allelic data does not contain information about the relative or absolute quantity of the individual alleles that may be present at that locus. Methods using non-allelic data only (that is, quantitative data from non-polymorphic alleles, or quantitative data from polymorphic loci but without regard to the allelic identity of each fragment) are referred to as quantitative methods. Typically, quantitative measurements are obtained for all possible alleles of the polymorphic locus of interest, with one value associated with the measured quantity for all of the alleles at that locus, in total. Non-allelic data for a polymorphic locus may be obtained by summing the quantitative allelic for each allele at that locus. When the allelic data is measured using high-throughput sequencing, the non-allelic data typically includes the number of reads of mapping to the locus of interest. The sequencing measurements could indicate the relative and/or absolute number of each of the alleles present at the locus, and the non-allelic data includes the sum of the reads, regardless of the allelic identity, mapping to the locus. In some embodiments the same set of sequencing measurements can be used to yield both allelic data and non-allelic

# US 11,530,454 B2

10

data. In some embodiments, the allelic data is used as part of a method to determine copy number at a chromosome of interest, and the produced non-allelic data can be used as part of a different method to determine copy number at a chromosome of interest. In some embodiments, the two methods are statistically orthogonal, and are combined to give a more accurate determination of the copy number at the chromosome of interest.

In some embodiments obtaining genetic data includes (i) acquiring DNA sequence information by laboratory techniques, e.g., by the use of an automated high throughput DNA sequencer, or (ii) acquiring information that had been previously obtained by laboratory techniques, wherein the information is electronically transmitted, e.g., by a computer over the internet or by electronic transfer from the sequencing device.

Additional exemplary sample preparation, amplification, and quantification methods are described in U.S. application Ser. No. 13/683,604, filed Nov. 21, 2012 (U.S. Publication 20 No. 2013/0123120 and U.S. Ser. No. 61/994,791, filed May 16, 2014, which is hereby incorporated by reference in its entirety). These methods can be used for analysis of any of the samples disclosed herein.

Exemplary Quantification Methods for Cell-Free DNA

If desired, that amount or concentration of cfDNA or cfRNA can be measured using standard methods. In some embodiments, the amount or concentration of cell-free mitochondrial DNA (cf mDNA) is determined. In some embodiments, the amount or concentration of cell-free DNA that 30 originated from nuclear DNA (cf nDNA) is determined. In some embodiments, the amount or concentration of cf mDNA and cf nDNA are determined simultaneously.

In some embodiments, qPCR is used to measure cf nDNA and/or cfm DNA (Kohler et al. "Levels of plasma circulating 35 cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors." Mol Cancer 8:105, 2009, 8:doi: 10.1186/1476-4598-8-105, which is hereby incorporated by reference in its entirety). For example, one or more loci from cf nDNA (such as Glyceraldehyd-3-phosphat-dehydroge- 40 nase, GAPDH) and one or more loci from cf mDNA (ATPase 8, MTATP 8) can be measured using multiplex qPCR. In some embodiments, fluorescence-labelled PCR is used to measure cf nDNA and/or cf mDNA (Schwarzenbach et al., "Evaluation of cell-free tumour DNA and RNA in 45 patients with breast cancer and benign breast disease." Mol Biosys 7:2848-2854, 2011, which is hereby incorporated by reference in its entirety). If desired, the normality distribution of the data can be determined using standard methods, such as the Shapiro-Wilk-Test. If desired, cf nDNA and 50 mDNA levels can be compared using standard methods, such as the Mann-Whitney-U-Test. In some embodiments, cf nDNA and/or mDNA levels are compared with other established prognostic factors using standard methods, such as the Mann-Whitney-U-Test or the Kruskal-Wallis-Test. Exemplary RNA Amplification, Quantification, and Analysis Methods

Any of the following exemplary methods may be used to amplify and optionally quantify RNA, such as such as cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, mitochondrial RNA, rRNA, or tRNA. In some embodiments, the miRNA is any of the miRNA molecules listed in the miRBase database available at the world wide web at mirbase.org, which is hereby incorporated by reference in its 65 entirety. Exemplary miRNA molecules include miR-509; miR-21, and miR-146a.

102

In some embodiments, reverse-transcriptase multiplex ligation-dependent probe amplification (RT-MLPA) is used to amplify RNA. In some embodiments, each set of hybridizing probes consists of two short synthetic oligonucleotides spanning the SNP and one long oligonucleotide (Li et al., Arch Gynecol Obstet. "Development of noninvasive prenatal diagnosis of trisomy 21 by RT-MLPA with a new set of SNP markers," Jul. 5, 2013, DOI 10.1007/s00404-013-2926-5. Schouten et al. "Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification." Nucleic Acids Res 30:e57, 2002; Deng et al. (2011) "Non-invasive prenatal diagnosis of trisomy 21 by reverse transcriptase multiplex ligation-dependent probe amplification," Clin, Chem. Lab Med. 49:641-646, 2011, which are each hereby incorporated by reference in its entirety).

In some embodiments, RNA is amplified with reverse-transcriptase PCR. In some embodiments, RNA is amplified with real-time reverse-transcriptase PCR, such as one-step real-time reverse-transcriptase PCR with SYBR GREEN I as previously described (Li et al., Arch Gynecol Obstet. "Development of noninvasive prenatal diagnosis of trisomy 21 by RT-MLPA with a new set of SNP markers," Jul. 5, 2013, DOI 10.1007/s00404-013-2926-5; Lo et al., "Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection," Nat Med 13:218-223, 2007; Tsui et al., Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. J Med Genet 41:461-467, 2004; Gu et al., J. Neurochem. 122:641-649, 2012, which are each hereby incorporated by reference in its entirety).

In some embodiments, a microarray is used to detect RNA. For example, a human miRNA microarray from Agilent Technologies can be used according to the manufacturer's protocol. Briefly, isolated RNA is dephosphorylated and ligated with pCp-Cy3. Labeled RNA is purified and hybridized to miRNA arrays containing probes for human mature miRNAs on the basis of Sanger miRBase release 14.0. The arrays is washed and scanned with use of a microarray scanner (G2565BA, Agilent Technologies). The intensity of each hybridization signal is evaluated by Agilent extraction software v9.5.3. The labeling, hybridization, and scanning may be performed according to the protocols in the Agilent miRNA microarray system (Gu et al., J. Neurochem. 122:641-649, 2012, which is hereby incorporated by reference in its entirety).

In some embodiments, a TaqMan assay is used to detect RNA. An exemplary assay is the TaqMan Array Human MicroRNA Panel v1.0 (Early Access) (Applied Biosystems), which contains 157 TaqMan MicroRNA Assays, including the respective reverse-transcription primers, PCR primers, and TaqMan probe (Chim et al., "Detection and characterization of placental microRNAs in maternal plasma," Clin Chem. 54(3):482-90, 2008, which is hereby incorporated by reference in its entirety).

If desired, the mRNA splicing pattern of one or more mRNAs can be determined using standard methods (Fackenthal1 and Godley, Disease Models & Mechanisms 1: 37-42, 2008, doi:10.1242/dmm.000331, which is hereby incorporated by reference in its entirety). For example, high-density microarrays and/or high-throughput DNA sequencing can be used to detect mRNA splice variants.

In some embodiments, whole transcriptome shotgun sequencing or an array is used to measure the transcriptome.

US 11,530,454 B2

103

**Exemplary Amplification Methods** 

Improved PCR amplification methods have also been developed that minimize or prevent interference due to the amplification of nearby or adjacent target loci in the same reaction volume (such as part of the sample multiplex PCR 5 reaction that simultaneously amplifies all the target loci). These methods can be used to simultaneously amplify nearby or adjacent target loci, which is faster and cheaper than having to separate nearby target loci into different reaction volumes so that they can be amplified separately to 10 avoid interference.

In some embodiments, the amplification of target loci is performed using a polymerase (e.g., a DNA polymerase, RNA polymerase, or reverse transcriptase) with low 5'→3' exonuclease and/or low strand displacement activity. In 15 some embodiments, the low level of 5'→3' exonuclease reduces or prevents the degradation of a nearby primer (e.g., an unextended primer or a primer that has had one or more nucleotides added to during primer extension). In some embodiments, the low level of strand displacement activity 20 reduces or prevents the displacement of a nearby primer (e.g., an unextended primer or a primer that has had one or more nucleotides added to it during primer extension). In some embodiments, target loci that are adjacent to each other (e.g., no bases between the target loci) or nearby (e.g., 25 loci are within 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 base) are amplified. In some embodiments, the 3' end of one locus is within 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 base of the 5' end of next downstream locus.

In some embodiments, at least 100, 200, 500, 750, 1,000; 30 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci are amplified, such as by the simultaneous amplification in one reaction volume. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products 3: are target amplicons. In various embodiments, the amount of amplified products that are target amplicons is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 98%, 90 to 99.5%, or 95 to 99.5%, inclusive. In some embodiments, at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% 40 of the targeted loci are amplified (e.g., amplified at least 5, 10, 20, 30, 50, or 100-fold compared to the amount prior to amplification), such as by the simultaneous amplification in one reaction volume. In various embodiments, the amount target loci that are amplified (e.g., amplified at least 5, 10, 45 20, 30, 50, or 100-fold compared to the amount prior to amplification) is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 99%, 90 to 99.5%, 95 to 99.9%, or 98 to 99.99% inclusive. In some embodiments, fewer non-target amplicons are produced, such as fewer amplicons 50 formed from a forward primer from a first primer pair and a reverse primer from a second primer pair. Such undesired non-target amplicons can be produced using prior amplification methods if, e.g., the reverse primer from the first primer pair and/or the forward primer from the second 55 primer pair are degraded and/or displaced.

In some embodiments, these methods allows longer extension times to be used since the polymerase bound to a primer being extended is less likely to degrade and/or displace a nearby primer (such as the next downstream 60 primer) given the low 5'→3' exonuclease and/or low strand displacement activity of the polymerase. In various embodiments, reaction conditions (such as the extension time and temperature) are used such that the extension rate of the polymerase allows the number of nucleotides that are added 65 to a primer being extended to be equal to or greater than 80, 90, 95, 100, 110, 120, 130, 140, 150, 175, or 200% of the

104

number of nucleotides between the 3' end of the primer binding site and the 5' end of the next downstream primer binding site on the same strand.

In some embodiments, a DNA polymerase is used produce DNA amplicons using DNA as a template. In some embodiments, a RNA polymerase is used produce RNA amplicons using DNA as a template. In some embodiments, a reverse transcriptase is used produce cDNA amplicons using RNA as a template.

In some embodiments, the low level of 5'→3' exonuclease of the polymerase is less than 80, 70, 60, 50, 40, 30, 20, 10, 5, 1, or 0.1% of the activity of the same amount of *Thermus aquaticus* polymerase ("Taq" polymerase, which is a commonly used DNA polymerase from a thermophilic bacterium, PDB 1BGX, EC 2.7.7.7, Murali et al., "Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme," Proc. Natl. Acad. Sci. USA 95:12562-12567, 1998, which is hereby incorporated by reference in its entirety) under the same conditions. In some embodiments, the low level of strand displacement activity of the polymerase is less than 80, 70, 60, 50, 40, 30, 20, 10, 5, 1, or 0.1% of the activity of the same amount of Taq polymerase under the same conditions.

In some embodiments, the polymerase is a PUSHION DNA polymerase, such as PHUSION High Fidelity DNA polymerase (M0530S, New England BioLabs, Inc.) or PHUSION Hot Start Flex DNA polymerase (M05355, New England BioLabs, Inc.; Frey and Suppman *BioChemica*. 2:34-35, 1995; Chester and Marshak *Analytical Biochemistry*. 209:284-290, 1993, which are each hereby incorporated by reference in its entirety). The PHUSION DNA polymerase is a *Pyrococcus*-like enzyme fused with a processivity-enhancing domain. PHUSION DNA polymerase possesses 5'→3' polymerase activity and 3'→5' exonuclease activity, and generates blunt-ended products. PHUSION DNA polymerase lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, the polymerase is a Q5® DNA Polymerase, such as Q5® High-Fidelity DNA Polymerase (M0491S, New England BioLabs, Inc.) or Q5® Hot Start High-Fidelity DNA Polymerase (M0493 S, New England BioLabs, Inc.). Q5® High-Fidelity DNA polymerase is a high-fidelity, thermostable, DNA polymerase with 3'->5' exonuclease activity, fused to a processivity-enhancing Sso7d domain. Q5® High-Fidelity DNA polymerase lacks 5'->3' exonuclease activity and strand displacement activity.

In some embodiments, the polymerase is a T4 DNA polymerase (M0203 S, New England BioLabs, Inc.; Tabor and Struh. (1989). "DNA-Dependent DNA Polymerases," In Ausebel et al. (Ed.), Current Protocols in Molecular Biology. 3.5.10-3.5.12. New York: John Wiley & Sons, Inc., 1989; Sambrook et al. Molecular Cloning: A Laboratory Manual. (2nd ed.), 5.44-5.47. Cold Spring Harbor. Cold Spring Harbor Laboratory Press, 1989, which are each hereby incorporated by reference in its entirety). T4 DNA Polymerase catalyzes the synthesis of DNA in the 5'→3' direction and requires the presence of template and primer. This enzyme has a 3'→5' exonuclease activity which is much more active than that found in DNA Polymerase I. T4 DNA polymerase lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, the polymerase is a *Sulfolobus* DNA Polymerase IV (M0327S, New England BioLabs, Inc.; (Boudsocq, et al. (2001). *Nucleic Acids Res.*, 29:4607-4616, 2001; McDonald, et al. (2006). *Nucleic Acids Res.*, 34:1102-1111, 2006, which are each hereby incorporated by reference

# US 11,530,454 B2

in its entirety). *Sulfolobus* DNA Polymerase IV is a thermostable Y-family lesion-bypass DNA Polymerase that efficiently synthesizes DNA across a variety of DNA template lesions McDonald, J. P. et al. (2006). *Nucleic Acids Res.*, 34, 1102-1111, which is hereby incorporated by reference in its entirety). *Sulfolobus* DNA Polymerase IV lacks 5'→3' exonuclease activity and strand displacement activity.

In some embodiments, if a primer binds a region with a SNP, the primer may bind and amplify the different alleles with different efficiencies or may only bind and amplify one 10 allele. For subjects who are heterozygous, one of the alleles may not be amplified by the primer. In some embodiments, a primer is designed for each allele. For example, if there are two alleles (e.g., a biallelic SNP), then two primers can be used to bind the same location of a target locus (e.g., a forward primer to bind the "A" allele and a forward primer to bind the "B" allele). Standard methods, such as the dbSNP database, can be used to determine the location of known SNPs, such as SNP hot spots that have a high heterozygosity rate.

In some embodiments, the amplicons are similar in size. In some embodiments, the range of the length of the target amplicons is less than 100, 75, 50, 25, 15, 10, or 5 nucleotides. In some embodiments (such as the amplification of target loci in fragmented DNA or RNA), the length of the 25 target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 and 75 nucleotides, inclusive. In some embodiments (such as the amplification of multiple target loci throughout an exon or gene), the length of the target amplicons is between 100 and 500 30 nucleotides, such as between 150 and 450 nucleotides, 200 and 400 nucleotides, 200 and 300 nucleotides, or 300 and 400 nucleotides, inclusive.

In some embodiments, multiple target loci are simultaneously amplified using a primer pair that includes a forward 35 and reverse primer for each target locus to be amplified in that reaction volume. In some embodiments, one round of PCR is performed with a single primer per target locus, and then a second round of PCR is performed with a primer pair per target locus. For example, the first round of PCR may be 40 performed with a single primer per target locus such that all the primers bind the same strand (such as using a forward primer for each target locus). This allows the PCR to amplify in a linear manner and reduces or eliminates amplification bias between amplicons due to sequence or length differences. In some embodiments, the amplicons are then amplified using a forward and reverse primer for each target locus. Exemplary Primer Design Methods

If desired, multiplex PCR may be performed using primers with a decreased likelihood of forming primer dimers. In particular, highly multiplexed PCR can often result in the production of a very high proportion of product DNA that results from unproductive side reactions such as primer dimer formation. In an embodiment, the particular primers that are most likely to cause unproductive side reactions may be removed from the primer library to give a primer library that will result in a greater proportion of amplified DNA that maps to the genome. The step of removing problematic primers, that is, those primers that are particularly likely to firm dimers has unexpectedly enabled extremely high PCR multiplexing levels for subsequent analysis by sequencing.

There are a number of ways to choose primers for a library where the amount of non-mapping primer dimer or other primer mischief products are minimized. Empirical data indicate that a small number of 'bad' primers are 65 responsible for a large amount of non-mapping primer dimer side reactions. Removing these 'bad' primers can increase

the percent of sequence reads that map to targeted loci. One way to identify the 'bad' primers is to look at the sequencing data of DNA that was amplified by targeted amplification; those primer dimers that are seen with greatest frequency can be removed to give a primer library that is significantly less likely to result in side product DNA that does not map to the genome. There are also publicly available programs that can calculate the binding energy of various primer combinations, and removing those with the highest binding energy will also give a primer library that is significantly less likely to result in side product DNA that does not map to the genome.

106

In some embodiments for selecting primers, an initial library of candidate primers is created by designing one or more primers or primer pairs to candidate target loci. A set of candidate target loci (such as SNPs) can selected based on publicly available information about desired parameters for the target loci, such as frequency of the SNPs within a target population or the heterozygosity rate of the SNPs. In one 20 embodiment, the PCR primers may be designed using the Primer3 program (the worldwide web at primer3. sourceforge.net; libprimer3 release 2.2.3, which is hereby incorporated by reference in its entirety). If desired, the primers can be designed to anneal within a particular annealing temperature range, have a particular range of GC contents, have a particular size range, produce target amplicons in a particular size range, and/or have other parameter characteristics. Starting with multiple primers or primer pairs per candidate target locus increases the likelihood that a primer or prime pair will remain in the library for most or all of the target loci. In one embodiment, the selection criteria may require that at least one primer pair per target locus remains in the library. That way, most or all of the target loci will be amplified when using the final primer library. This is desirable for applications such as screening for deletions or duplications at a large number of locations in the genome or screening for a large number of sequences (such as polymorphisms or other mutations) associated with a disease or an increased risk for a disease. If a primer pair from the library would produces a target amplicon that overlaps with a target amplicon produced by another primer pair, one of the primer pairs may be removed from the library to prevent interference.

In some embodiments, an "undesirability score" (higher score representing least desirability) is calculated (such as calculation on a computer) for most or all of the possible combinations of two primers from a library of candidate primers. In various embodiments, an undesirability score is calculated for at least 80, 90, 95, 98, 99, or 99.5% of the possible combinations of candidate primers in the library. Each undesirability score is based at least in part on the likelihood of dimer formation between the two candidate primers. If desired, the undesirability score may also be based on one or more other parameters selected from the group consisting of heterozygosity rate of the target locus, disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, specificity of the candidate primer for the target locus, size of the candidate primer, melting temperature of the target amplicon, GC content of the target amplicon, amplification efficiency of the target amplicon, size of the target amplicon, and distance from the center of a recombination hotspot. In some embodiments, the specificity of the candidate primer for the target locus includes the likelihood that the candidate primer will mis-prime by binding and amplifying a locus other than the target locus it was designed to amplify. In

# US 11,530,454 B2

some embodiments, one or more or all the candidate primers that mis-prime are removed from the library. In some embodiments to increase the number of candidate primers to choose from, candidate primers that may mis-prime are not removed from the library. If multiple factors are considered, the undesirability score may be calculated based on a weighted average of the various parameters. The parameters may be assigned different weights based on their importance for the particular application that the primers will be used for. In some embodiments, the primer with the highest 10 undesirability score is removed from the library. If the removed primer is a member of a primer pair that hybridizes to one target locus, then the other member of the primer pair may be removed from the library. The process of removing primers may be repeated as desired. In some embodiments, 15 the selection method is performed until the undesirability scores for the candidate primer combinations remaining in the library are all equal to or below a minimum threshold. In some embodiments, the selection method is performed until the number of candidate primers remaining in the library is 20 reduced to a desired number.

107

In various embodiments, after the undesirability scores are calculated, the candidate primer that is part of the greatest number of combinations of two candidate primers with an undesirability score above a first minimum threshold 25 is removed from the library. This step ignores interactions equal to or below the first minimum threshold since these interactions are less significant. If the removed primer is a member of a primer pair that hybridizes to one target locus, then the other member of the primer pair may be removed 30 from the library. The process of removing primers may be repeated as desired. In some embodiments, the selection method is performed until the undesirability scores for the candidate primer combinations remaining in the library are all equal to or below the first minimum threshold. If the 35 number of candidate primers remaining in the library is higher than desired, the number of primers may be reduced by decreasing the first minimum threshold to a lower second minimum threshold and repeating the process of removing primers. If the number of candidate primers remaining in the 4 library is lower than desired, the method can be continued by increasing the first minimum threshold to a higher second minimum threshold and repeating the process of removing primers using the original candidate primer library, thereby allowing more of the candidate primers to remain in the 45 library. In some embodiments, the selection method is performed until the undesirability scores for the candidate primer combinations remaining in the library are all equal to or below the second minimum threshold, or until the number of candidate primers remaining in the library is reduced to 50 a desired number.

If desired, primer pairs that produce a target amplicon that overlaps with a target amplicon produced by another primer pair can be divided into separate amplification reactions. Multiple PCR amplification reactions may be desirable for 55 applications in which it is desirable to analyze all of the candidate target loci (instead of omitting candidate target loci from the analysis due to overlapping target amplicons).

These selection methods minimize the number of candidate primers that have to be removed from the library to 60 achieve the desired reduction in primer dimers. By removing a smaller number of candidate primers from the library, more (or all) of the target loci can be amplified using the resulting primer library.

Multiplexing large numbers of primers imposes consid-65 erable constraint on the assays that can be included. Assays that unintentionally interact result in spurious amplification

further constraints. In an embodiment, it is possible to begin with a very large number of potential SNP targets (between about 500 to greater than 1 million) and attempt to design primers to amplify each SNP. Where primers can be designed it is possible to attempt to identify primer pairs likely to form spurious products by evaluating the likelihood of spurious primer duplex formation between all possible pairs of primers using published thermodynamic parameters for DNA duplex formation. Primer interactions may be ranked by a scoring function related to the interaction and primers with the worst interaction scores are eliminated until the number of primers desired is met. In cases where SNPs likely to be heterozygous are most useful, it is possible to also rank the list of assays and select the most heterozygous compatible assays. Experiments have validated that primers with high interaction scores are most likely to form primer dimers. At high multiplexing it is not possible to eliminate all spurious interactions, but it is essential to remove the primers or pairs of primers with the highest interaction scores in silico as they can dominate an entire reaction, greatly limiting amplification from intended targets. We have performed this procedure to create multiplex primer

sets of up to and in some cases more than 10,000 primers.

The improvement due to this procedure is substantial, enabling amplification of more than 80%, more than 90%,

more than 95%, more than 98%, and even more than 99% on

target products as determined by sequencing of all PCR

products, as compared to 10% from a reaction in which the

worst primers were not removed. When combined with a

108

products. The size constraints of miniPCR may result in

partial semi-nested approach as previously described, more than 90%, and even more than 95% of amplicons may map to the targeted sequences.

Note that there are other methods for determining which PCR probes are likely to form dimers. In an embodiment, analysis of a pool of DNA that has been amplified using a non-optimized set of primers may be sufficient to determine problematic primers. For example, analysis may be done using sequencing, and those dimers which are present in the greatest number are determined to be those most likely to form dimers, and may be removed. In an embodiment, the

method of primer design may be used in combination with

the mini-PCR method described herein.

The use of tags on the primers may reduce amplification and sequencing of primer dimer products. In some embodiments, the primer contains an internal region that forms a loop structure with a tag. In particular embodiments, the primers include a 5' region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In some embodiments, the loop region may lie between two binding regions where the two binding regions are designed to bind to contiguous or neighboring regions of template DNA. In various embodiments, the length of the 3' region is at least 7 nucleotides. In some embodiments, the length of the 3' region is between 7 and 20 nucleotides, such as between 7 to 15 nucleotides, or 7 to 10 nucleotides, inclusive. In various embodiments, the primers include a 5 region that is not specific for a target locus (such as a tag or a universal primer binding site) followed by a region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. Tag-primers can be used to shorten necessary target-specific sequences to below 20, below 15, below 12, and even below 10 base pairs. This can be serendipitous with standard primer design when the target sequence is fragmented within the primer

# US 11,530,454 B2

109

binding site or, or it can be designed into the primer design. Advantages of this method include: it increases the number of assays that can be designed for a certain maximal amplicon length, and it shortens the "non-informative" sequencing of primer sequence. It may also be used in combination 5 with internal tagging.

In an embodiment, the relative amount of nonproductive products in the multiplexed targeted PCR amplification can be reduced by raising the annealing temperature. In cases where one is amplifying libraries with the same tag as the 10 target specific primers, the annealing temperature can be increased in comparison to the genomic DNA as the tags will contribute to the primer binding. In some embodiments reduced primer concentrations are used, optionally along with longer annealing times. In some embodiments the 15 annealing times may be longer than 3 minutes, longer than 5 minutes, longer than 8 minutes, longer than 10 minutes, longer than 15 minutes, longer than 20 minutes, longer than 30 minutes, longer than 60 minutes, longer than 120 minutes, longer than 240 minutes, longer than 480 minutes, and 20 even longer than 960 minutes. In certain illustrative embodiments, longer annealing times are used along with reduced primer concentrations. In various embodiments, longer than normal extension times are used, such as greater than 3, 5, 8, 10, or 15 minutes. In some embodiments, the primer 25 concentrations are as low as 50 nM, 20 nM, 10 nM, 5 nM, 1 nM, and lower than 1 nM. This surprisingly results in robust performance for highly multiplexed reactions, for example 1,000-plex reactions, 2,000-plex reactions, 5,000plex reactions, 10,000-plex reactions, 20,000-plex reactions, 30 50,000-plex reactions, and even 100,000-plex reactions. In an embodiment, the amplification uses one, two, three, four or five cycles run with long annealing times, followed by PCR cycles with more usual annealing times with tagged primers

To select target locations, one may start with a pool of candidate primer pair designs and create a thermodynamic model of potentially adverse interactions between primer pairs, and then use the model to eliminate designs that are incompatible with other the designs in the pool.

In an embodiment, the invention features a method of decreasing the number of target loci (such as loci that may contain a polymorphism or mutation associated with a disease or disorder or an increased risk for a disease or disorder such as cancer) and/or increasing the disease load 45 that is detected (e.g., increasing the number of polymorphisms or mutations that are detected). In some embodiments, the method includes ranking (such as ranking from highest to lowest) loci by frequency or reoccurrence of a polymorphism or mutation (such as a single nucleotide 50 variation, insertion, or deletion, or any of the other variations described herein) in each locus among subjects with the disease or disorder such as cancer. In some embodiments, PCR primers are designed to some or all of the loci. During selection of PCR primers for a library of primers, primers to 55 loci that have a higher frequency or reoccurrence (higher ranking loci) are favored over those with a lower frequency or reoccurrence (lower ranking loci). In some embodiments, this parameter is included as one of the parameters in the calculation of the undesirability scores described herein. If 60 desired, primers (such as primers to high ranking loci) that are incompatible with other designs in the library can be included in a different PCR library/pool. In some embodiments, multiple libraries/pools (such as 2, 3, 4, 5 or more) are used in separate PCR reactions to enable amplification of 65 all (or a majority) of the loci represented by all the libraries/ pools. In some embodiment, this method is continued until

110

sufficient primers are included in one or more libraries/pools such that the primers, in aggregate, enable the desired disease load to be captured for the disease or disorder (e.g., such as by detection of at least 80, 85, 90, 95, or 99% of the disease load).

**Exemplary Primer Libraries** 

In one aspect, the invention features libraries of primers, such as primers selected from a library of candidate primers using any of the methods of the invention. In some embodiments, the library includes primers that simultaneously hybridize (or are capable of simultaneously hybridizing) to or that simultaneously amplify (or are capable of simultaneously amplifying) at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci in one reaction volume. In various embodiments, the library includes primers that simultaneously amplify (or are capable of simultaneously amplifying) between 100 to 500; 500 to 1,000; 1,000 to 2,000; 2,000 to 5,000; 5,000 to 7,500; 7,500 to 10,000; 10,000 to 20,000; 20,000 to 25,000; 25,000 to 30,000; 30,000 to 40,000; 40,000 to 50,000; 50,000 to 75,000; or 75,000 to 100,000 different target loci in one reaction volume, inclusive. In various embodiments, the library includes primers that simultaneously amplify (or are capable of simultaneously amplifying) between 1,000 to 100,000 different target loci in one reaction volume, such as between 1,000 to 50,000; 1,000 to 30,000; 1,000 to 20,000; 1,000 to 10,000; 2,000 to 30,000; 2,000 to 20,000; 2,000 to 10,000; 5,000 to 30,000; 5,000 to 20,000; or 5,000 to 10,000 different target loci, inclusive. In some embodiments, the library includes primers that simultaneously amplify (or are capable of simultaneously amplifying) the target loci in one reaction volume such that less than 60, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0.1, or 0.5% of the amplified products are primer dimers. The various embodiments, the amount of amplified products that are primer dimers is between 0.5 to 60%, such as between 0.1 to 40%, 0.1 to 20%, 0.25 to 20%, 0.25 to 10%, 0.5 to 20%, 0.5 to 10%, 1 to 20%, or 1 to 10%, inclusive. In some embodiments, the primers simultaneously amplify (or are capable of simultaneously amplifying) the target loci in one reaction volume such that at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the amplified products are target amplicons. In various embodiments, the amount of amplified products that are target amplicons is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 98%, 90 to 99.5%, or 95 to 99.5%, inclusive. In some embodiments, the primers simultaneously amplify (or are capable of simultaneously amplifying) the target loci in one reaction volume such that at least 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 99.5% of the targeted loci are amplified (e.g., amplified at least 5, 10, 20, 30, 50, or 100-fold compared to the amount prior to amplification). In various embodiments, the amount target loci that are amplified (e.g., amplified at least 5, 10, 20, 30, 50, or 100-fold compared to the amount prior to amplification) is between 50 to 99.5%, such as between 60 to 99%, 70 to 98%, 80 to 99%, 90 to 99.5%, 95 to 99.9%, or 98 to 99.99% inclusive. In some embodiments, the library of primers includes at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 primer pairs, wherein each pair of primers includes a forward test primer and a reverse test primer where each pair of test primers hybridize to a target locus. In some embodiments, the library of primers includes at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 individual primers that

## US 11,530,454 B2

11

each hybridize to a different target locus, wherein the individual primers are not part of primer pairs.

In various embodiments, the concentration of each primer is less than 100, 75, 50, 25, 20, 10, 5, 2, or 1 nM, or less than 500, 100, 10, or 1 uM. In various embodiments, the concentration of each primer is between 1 uM to 100 nM, such as between 1 uM to 1 nM, 1 to 75 nM, 2 to 50 nM or 5 to 50 nM, inclusive. In various embodiments, the GC content of the primers is between 30 to 80%, such as between 40 to 70%, or 50 to 60%, inclusive. In some embodiments, the 10 range of GC content of the primers is less than 30, 20, 10, or 5%. In some embodiments, the range of GC content of the primers is between 5 to 30%, such as 5 to 20% or 5 to 10%, inclusive. In some embodiments, the melting temperature  $(T_m)$  of the test primers is between 40 to 80° C., such as 50 15 to 70° C., 55 to 65° C., or 57 to 60.5° C., inclusive. In some embodiments, the  $T_m$  is calculated using the Primer3 program (libprimer3 release 2.2.3) using the built-in SantaLucia parameters (the world wide web at primer3.sourceforge.net). In some embodiments, the range of melting temperature of 20 the primers is less than 15, 10, 5, 3, or 1° C. In some embodiments, the range of melting temperature of the primers is between 1 to 15° C., such as between 1 to 10° C., 1 to 5° C., or 1 to 3° C., inclusive. In some embodiments, the length of the primers is between 15 to 100 nucleotides, such 25 as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, or 20 to 65 nucleotides, inclusive. In some embodiments, the range of the length of the primers is less than 50, 40, 30, 20, 10, or 5 nucleotides. In some embodiments, the range of the length of the primers 30 is between 5 to 50 nucleotides, such as 5 to 40 nucleotides, 5 to 20 nucleotides, or 5 to 10 nucleotides, inclusive. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodi- 35 ments, the range of the length of the target amplicons is less than 50, 25, 15, 10, or 5 nucleotides. In some embodiments, the range of the length of the target amplicons is between 5 to 50 nucleotides, such as 5 to 25 nucleotides, 5 to 15 nucleotides, or 5 to 10 nucleotides, inclusive. In some 40 embodiments, the library does not comprise a microarray. In some embodiments, the library comprises a microarray.

In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include one or more linkages between adjacent nucleotides other than a natu- 45 rally-occurring phosphodiester linkage. Examples of such linkages include phosphoramide, phosphorothioate, and phosphorodithioate linkages. In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include a thiophosphate (such as a monothiophos- 50 phate) between the last 3' nucleotide and the second to last 3' nucleotide. In some embodiments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include a thiophosphate (such as a monothiophosphate) between the last 2, 3, 4, or 5 nucleotides at the 3' end. In some embodi- 55 ments, some (such as at least 80, 90, or 95%) or all of the adaptors or primers include a thiophosphate (such as a monothiophosphate) between at least 1, 2, 3, 4, or 5 nucleotides out of the last 10 nucleotides at the 3' end. In some embodiments, such primers are less likely to be cleaved or 60 degraded. In some embodiments, the primers do not contain an enzyme cleavage site (such as a protease cleavage site).

Additional exemplary multiplex PCR methods and libraries are described in U.S. application Ser. No. 13/683,604, filed Nov. 21, 2012 (U.S. Publication No. 2013/0123120) 65 and U.S. Ser. No. 61/994,791, filed May 16, 2014, which are each hereby incorporated by reference in its entirety). These

112

methods and libraries can be used for analysis of any of the samples disclosed herein and for use in any of the methods of the invention.

Exemplary Primer Libraries for Detection of Recombination In some embodiments, primers in the primer library are designed to determine whether or not recombination occurred at one or more known recombination hotspots (such as crossovers between homologous human chromosomes). Knowing what crossovers occurred between chromosomes allows more accurate phased genetic data to be determined for an individual. Recombination hotspots are local regions of chromosomes in which recombination events tend to be concentrated. Often they are flanked by "coldspots," regions of lower than average frequency of recombination. Recombination hotspots tend to share a similar morphology and are approximately 1 to 2 kb in length. The hotspot distribution is positively correlated with GC content and repetitive element distribution. A partially degenerated 13-mer motif CCNCCNTNNCCNC plays a role in some hotspot activity. It has been shown that the zinc finger protein called PRDM9 binds to this motif and initiates recombination at its location. The average distance between the centers of recombination hot spots is reported to be ~80 kb. In some embodiments, the distance between the centers of recombination hot spots ranges between ~3 kb to ~100 kb. Public databases include a large number of known human recombination hotspots, such as the HUMHOT and International HapMap Project databases (see, for example, Nishant et al., "HUMHOT: a database of human meiotic recombination hot spots," Nucleic Acids Research, 34: D25-D28, 2006, Database issue; Mackiewicz et al., "Distribution of Recombination Hotspots in the Human Genome-A Comparison of Computer Simulations with Real Data" PLoS ONE 8(6): e65272, doi:10.1371/journal.pone. 0065272; and the world wide web at hapmap.ncbi.nlm.nih. gov/downloads/index.html.en, which are each hereby incorporated by reference in its entirety).

In some embodiments, primers in the primer library are clustered at or near recombination hotspots (such as known human recombination hotspots). In some embodiments, the corresponding amplicons are used to determine the sequence within or near a recombination hotspot to determine whether or not recombination occurred at that particular hotspot (such as whether the sequence of the amplicon is the sequence expected if a recombination had occurred or the sequence expected if a recombination had not occurred). In some embodiments, primers are designed to amplify part or all of a recombination hotspot (and optionally sequence flanking a recombination hotspot). In some embodiments, long read sequencing (such as sequencing using the Moleculo Technology developed by Illumina to sequence up to ~10 kb) or paired end sequencing is used to sequence part or all of a recombination hotspot. Knowledge of whether or not a recombination event occurred can be used to determine which haplotype blocks flank the hotspot. If desired, the presence of particular haplotype blocks can be confirmed using primers specific to regions within the haplotype blocks. In some embodiments, it is assumed there are no crossovers between known recombination hotspots. In some embodiments, primers in the primer library are clustered at or near the ends of chromosomes. For example, such primers can be used to determine whether or not a particular arm or section at the end of a chromosome is present. In some embodiments, primers in the primer library are clustered at or near recombination hotspots and at or near the ends of chromosomes.

## US 11,530,454 B2

113

In some embodiments, the primer library includes one or more primers (such as at least 5; 10; 50; 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; or 50,000 different primers or different primer pairs) that are specific for a recombination hotspot (such as a known human recombination hotspot) and/or are specific for a region near a recombination hotspot (such as within 10, 8, 5, 3, 2, 1, or 0.5 kb of the 5' or 3' end of a recombination hotspot). In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific 10 for the same recombination hotspot, or are specific for the same recombination hotspot or a region near the recombination hotspot. In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for a region between recombination hotspots (such 15 as a region unlikely to have undergone recombination); these primers can be used to confirm the presence of haplotype blocks (such as those that would be expected depending on whether or not recombination has occurred). In some embodiments, at least 10, 20, 30, 40, 50, 60, 70, 80, 20 or 90% of the primers in the primer library are specific for a recombination hotspot and/or are specific for a region near a recombination hotspot (such as within 10, 8, 5, 3, 2, 1, or 0.5 kb of the 5' or 3' end of the recombination hotspot). In some embodiments, the primer library is used to determine 25 whether or not recombination has occurred at greater than or equal to 5; 10; 50; 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; or 50,000 different recombination hotspots (such as known human recombination hotspots). In some embodiments, the regions 30 targeted by primers to a recombination hotspot or nearby region are approximately evenly spread out along that portion of the genome. In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for the a region at or near the end of a 3: chromosome (such as a region within 20, 10, 5, 1, 0.5, 0.1, 0.01, or 0.001 mb from the end of a chromosome). In some embodiments, at least 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the primers in the primer library are specific for the a region at or near the end of a chromosome (such as a region 40 within 20, 10, 5, 1, 0.5, 0.1, 0.01, or 0.001 mb from the end of a chromosome). In some embodiments, at least 1, 5, 10, 20, 40, 60, 80, 100, or 150 different primer (or primer pairs) are specific for a region within a potential microdeletion in a chromosome. In some embodiments, at least 10, 20, 30, 40, 45 50, 60, 70, 80, or 90% of the primers in the primer library are specific for a region within a potential microdeletion in a chromosome. In some embodiments, at least 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the primers in the primer library are specific for a recombination hotspot, a region near a 50 recombination hotspot, a region at or near the end of a chromosome, or a region within a potential microdeletion in a chromosome.

#### Exemplary Kits

In one aspect, the invention features a kit, such as a kit for 55 amplifying target loci in a nucleic acid sample for detecting deletions and/or duplications of chromosome segments or entire chromosomes using any of the methods described herein). In some embodiments, the kit can include any of the primer libraries of the invention. In an embodiment, the kit comprises a plurality of inner forward primers and optionally outer forward primers and outer reverse primers, where each of the primers is designed to hybridize to the region of DNA immediately upstream and/or downstream from one of the target sites (e.g., polymorphic sites) on the target chromosome(s) or chromosome segment(s), and optionally addi-

114

tional chromosomes or chromosome segments. In some embodiments, the kit includes instructions for using the primer library to amplify the target loci, such as for detecting one or more deletions and/or duplications of one or more chromosome segments or entire chromosomes using any of the methods described herein.

In certain embodiments, kits of the invention provide primer pairs for detecting chromosomal aneuploidy and CNV determination, such as primer pairs for massively multiplex reactions for detecting chromosomal aneuploidy such as CNV (CoNVERGe) (Copy Number Variant Events Revealed Genotypically) and/or SNVs. In these embodiments, the kits can include between at least 100, 200, 250, 300, 500, 1000, 2000, 2500, 3000, 5000, 10,000, 20,000, 25,000, 28,000, 50,000, or 75,000 and at most 200, 250, 300, 500, 1000, 2000, 2500, 3000, 5000, 10,000, 20,000, 25,000, 28,000, 50,000, 75,000, or 100,000 primer pairs that are shipped together. The primer pairs can be contained in a single vessel, such as a single tube or box, or multiple tubes or boxes. In certain embodiments, the primer pairs are pre-qualified by a commercial provider and sold together, and in other embodiments, a customer selects custom gene targets and/or primers and a commercial provider makes and ships the primer pool to the customer neither in one tube or a plurality of tubes. In certain exemplary embodiments, the kits include primers for detecting both CNVs and SNVs, especially CNVs and SNVs known to be correlated to at least one type of cancer.

Kits for circulating DNA detection according to some embodiments of the present invention, include standards and/or controls for circulating DNA detection. For example, in certain embodiments, the standards and/or controls are sold and optionally shipped and packaged together with primers used to perform the amplification reactions provided herein, such as primers for performing CoNVERGe. In certain embodiments, the controls include polynucleotides such as DNA, including isolated genomic DNA that exhibits one or more chromosomal aneuploidies such as CNV and/or includes one or more SNVs. In certain embodiments, the standards and/or controls are called PlasmArt standards and include polynucleotides having sequence identity to regions of the genome known to exhibit CNV, especially in certain inherited diseases, and in certain disease states such as cancer, as well as a size distribution that reflects that of cfDNA fragments naturally found in plasma. Exemplary methods for making PlasmArt standards are provided in the examples herein. In general, genomic DNA from a source known to include a chromosomal aneuoploidy is isolated, fragmented, purified and size selected.

Accordingly, artificial cfDNA polynucleotide standards and/or controls can be made by spiking isolated polynucleotide samples prepared as summarized above, into DNA samples known not to exhibit a chromosomal aneuploidy and/or SNVs, at concentrations similar to those observed for cfDNA in vivo, such as between, for example, 0.01% and 20%, 0.1 and 15%, or 0.4 and 10% of DNA in that fluid. These standards/controls can be used as controls for assay design, characterization, development, and/or validation, and as quality control standards during testing, such as cancer testing performed in a CLIA lab and/or as standards included in research use only or diagnostic test kits.

Exemplary Normalization/Correction Methods

In some embodiments, measurements for different loci, chromosome segments, or chromosomes are adjusted for bias, such as bias due to differences in GC content or bias due to other differences in amplification efficiency or adjusted for sequencing errors. In some embodiments, mea-

## US 11,530,454 B2

115

surements for different alleles for the same locus are adjusted for differences in metabolism, apoptosis, histones, inactivation, and/or amplification between the alleles. In some embodiments, measurements for different alleles for the same locus in RNA are adjusted for differences in transcription rates or stability between different RNA alleles. Exemplary Methods for Phasing Genetic Data

In some embodiments, genetic data is phased using the methods described herein or any known method for phasing genetic data (see, e.g., PCT Publ. No. WO2009/105531, 10 filed Feb. 9, 2009, and PCT Publ. No. WO2010/017214, filed Aug. 4, 2009; U.S. Publ. No. 2013/0123120, Nov. 21, 2012; U.S. Publ. No. 2011/0033862, filed Oct. 7, 2010; U.S. Publ. No. 2011/0033862, filed Aug. 19, 2010; U.S. Publ. No. 2011/0178719, filed Feb. 3, 2011; U.S. Pat. No. 8,515,679, filed Mar. 17, 2008; U.S. Publ. No. 2007/0184467, filed Nov. 22, 2006; U.S. Publ. No. 2008/0243398, filed Mar. 17, 2008, and U.S. Ser. No. 61/994,791, filed May 16, 2014, which are each hereby incorporated by reference in its entirety). In some embodiments, the phase is determined for 20 one or more regions that are known or suspected to contain a CNV of interest. In some embodiments, the phase is also determined for one or more regions flanking the CNV region(s) and/or for one or more reference regions. In one embodiment, genetic data of an individual (e.g., an indi- 25 vidual being tested using the methods of the invention or a relative of a gestating fetus or embryo, such as a parent of the fetus or embryo) is phased by inference by measuring tissue from the individual that is haploid, for example by measuring one or more sperm or eggs. In one embodiment, 30 an individual's genetic data is phased by inference using the measured genotypic data of one or more first degree relatives, such as the individual's parents (e.g., sperm from the individual's father) or siblings.

In one embodiment, an individual's genetic data is phased 35 by dilution where the DNA or RNA is diluted in one or a plurality of wells, such as by using digital PCR. In some embodiments, the DNA or RNA is diluted to the point where there is expected to be no more than approximately one copy of each haplotype in each well, and then the DNA or RNA 40 in the one or more wells is measured. In some embodiments, cells are arrested at phase of mitosis when chromosomes are tight bundles, and microfluidics is used to put separate chromosomes in separate wells. Because the DNA or RNA is diluted, it is unlikely that more than one haplotype is in the 4. same fraction (or tube). Thus, there may be effectively a single molecule of DNA in the tube, which allows the haplotype on a single DNA or RNA molecule to be determined. In some embodiments, the method includes dividing a DNA or RNA sample into a plurality of fractions such that 50 at least one of the fractions includes one chromosome or one chromosome segment from a pair of chromosomes, and genotyping (e.g., determining the presence of two or more polymorphic loci) the DNA or RNA sample in at least one of the fractions, thereby determining a haplotype. In some 55 embodiments, the genotyping involves sequencing (such as shotgun sequencing or single molecule sequencing), a SNP array to detect polymorphic loci, or multiplex PCR. In some embodiments, the genotyping involves use of a SNP array to detect polymorphic loci, such as at least 100; 200; 500; 750; 60 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci. In some embodiments, the genotyping involves the use of multiplex PCR. In some embodiments, the method involves contacting the sample in a fraction with a library of 65 primers that simultaneously hybridize to at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000;

116

25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci (such as SNPs) to produce a reaction mixture; and subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that are measured with a high throughput sequencer to produce sequencing data. In some embodiments, RNA (such as mRNA) is sequenced. Since mRNA contains only exons, sequencing mRNA allows alleles to be determined for polymorphic loci (such as SNPs) over a large distance in the genome, such as a few megabases. In some embodiments, a haplotype of an individual is determined by chromosome sorting. An exemplary chromosome sorting method includes arresting cells at phase of mitosis when chromosomes are tight bundles and using microfluidics to put separate chromosomes in separate wells. Another method involves collecting single chromosomes using FACS-mediated single chromosome sorting. Standard methods (such as sequencing or an array) can be used to identify the alleles on a single chromosome to determine a haplotype of the individual.

In some embodiments, a haplotype of an individual is determined by long read sequencing, such as by using the Moleculo Technology developed by Illumina. In some embodiments, the library prep step involves shearing DNA into fragments, such as fragments of ~10 kb size, diluting the fragments and placing them into wells (such that about 3,000 fragments are in a single well), amplifying fragments in each well by long-range PCR and cutting into short fragments and barcoding the fragments, and pooling the barcoded fragments from each well together to sequence them all. After sequencing, the computational steps involve separating the reads from each well based on the attached barcodes and grouping them into fragments, assembling the fragments at their overlapping heterozygous SNVs into haplotype blocks, and phasing the blocks statistically based on a phased reference panel and producing long haplotype contigs.

In some embodiments, a haplotype of the individual is determined using data from a relative of the individual. In some embodiments, a SNP array is used to determine the presence of at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci in a DNA or RNA sample from the individual and a relative of the individual. In some embodiments, the method involves contacting a DNA sample from the individual and/or a relative of the individual with a library of primers that simultaneously hybridize to at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci (such as SNPs) to produce a reaction mixture; and subjecting the reaction mixture to primer extension reaction conditions to produce amplified products that are measured with a high throughput sequencer to produce sequencing

In one embodiment, an individual's genetic data is phased using a computer program that uses population based haplotype frequencies to infer the most likely phase, such as HapMap-based phasing. For example, haploid data sets can be deduced directly from diploid data using statistical methods that utilize known haplotype blocks in the general population (such as those created for the public HapMap Project and for the Perlegen Human Haplotype Project). A haplotype block is essentially a series of correlated alleles that occur repeatedly in a variety of populations. Since these haplotype blocks are often ancient and common, they may be used to predict haplotypes from diploid genotypes. Publicly available algorithms that accomplish this task include an imperfect phylogeny approach, Bayesian approaches

## US 11,530,454 B2

117

based on conjugate priors, and priors from population genetics. Some of these algorithms use a hidden Markov model.

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from genotype data, such as an algorithm that uses localized haplotype 5 clustering (see, e.g., Browning and Browning, "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering" Am J Hum Genet. November 2007; 81(5): 1084-1097, which is hereby incorporated by reference in its entirety). An exemplary program is Beagle version: 3.3.2 or version 4 (available at the world wide web at hfaculty,washington.edu/browning/beagle/beagle.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from genotype data, such as an algorithm that uses the decay of linkage disequilibrium with distance, the order and spacing of genotyped markers, missing-data imputation, recombination rate estimates, or a combination thereof (see, e.g., Stephens and Scheet, "Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation" Am. J. Hum. Genet. 76:449-462, 2005, which is hereby incorporated by reference in its entirety). An exemplary program is PHASE v.2.1 or v2.1.1. (available at the world wide web at stephenslab.uchicago.edu/software.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from population genotype data, such as an algorithm that allows cluster memberships to change continuously along the chromosome according to a hidden Markov model. This approach is flexible, allowing for both "block-like" patterns of linkage disequilibrium and gradual decline in linkage disequilibrium with distance (see, e.g., Scheet and Stephens, "A fast and 35 flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase." Am J Hum Genet, 78:629-644, 2006, which is hereby incorporated by reference in its entirety). An exemplary program is fastPHASE (available at the world wide web at stephenslab.uchicago.edu/software.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using a genotype imputation method, such as a method that uses one or more of the following reference datasets: Hap- 45 Map dataset, datasets of controls genotyped on multiple SNP chips, and densely typed samples from the 1,000 Genomes Project. An exemplary approach is a flexible modelling framework that increases accuracy and combines information across multiple reference panels (see, e.g., 50 Howie, Donnelly, and Marchini (2009) "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS Genetics 5(6): e1000529, 2009, which is hereby incorporated by reference in its entirety). Exemplary programs are IMPUTE 55 or IMPUTE version 2 (also known as IMPUTE2) (available at the world wide web at mathgen.stats.ox.ac.uk/impute/ impute\_v2.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased 60 using an algorithm that infers haplotypes, such as an algorithm that infers haplotypes under the genetic model of coalescence with recombination, such as that developed by Stephens in PHASE v2.1. The major algorithmic improvements rely on the use of binary trees to represent the sets of 65 candidate haplotypes for each individual. These binary tree representations: (1) speed up the computations of posterior

118

probabilities of the haplotypes by avoiding the redundant operations made in PHASE v2.1, and (2) overcome the exponential aspect of the haplotypes inference problem by the smart exploration of the most plausible pathways (i.e., haplotypes) in the binary trees (see, e.g., Delaneau, Coulonges and Zagury, "Shape-IT: new rapid and accurate algorithm for haplotype inference," BMC Bioinformatics 9:540, 2008 doi:10.1186/1471-2105-9-540, which is hereby incorporated by reference in its entirety). An exemplary program is SHAPEIT (available at the world wide web at mathgen. stats.ox.ac.uk/genetics software/shapeit/shapeit.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from population genotype data, such as an algorithm that uses haplotype-fragment frequencies to obtain empirically based probabilities for longer haplotypes. In some embodiments, the algorithm reconstructs haplotypes so that they have maximal local coherence (see, e.g., Eronen, Geerts, and Toivonen, "HaploRec: Efficient and accurate large-scale reconstruction of haplotypes," *BMC Bioinformatics* 7:542, 2006, which is hereby incorporated by reference in its entirety). An exemplary program is HaploRec, such as HaploRec version 2.3. (available at the world wide web at cs.helsinki.fi/group/genetics/haplotyping.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from population genotype data, such as an algorithm that uses a partition-ligation strategy and an expectation-maximization-based algorithm (see, e.g., Qin, Niu, and Liu, "Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms," Am J Hum Genet. 71(5): 1242-1247, 2002, which is hereby incorporated by reference in its entirety). An exemplary program is PL-EM (available at the world wide web at people.fas.harvard.edu/~junliu/plem/click.html, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from population genotype data, such as an algorithm for simultaneously phasing genotypes into haplotypes and block partitioning. In some embodiments, an expectation-maximization algorithm is used (see, e.g., Kimmel and Shamir, "GERBIL: Genotype Resolution and Block Identification Using Likelihood," Proceedings of the National Academy of Sciences of the United States of America (PNAS) 102: 158-162, 2005, which is hereby incorporated by reference in its entirety). An exemplary program is GERBIL, which is available as part of the GEVALT version 2 program (available at the world wide web at acgt.cs.tau.ac.il/gevalt/, which is hereby incorporated by reference in its entirety).

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from population genotype data, such as an algorithm that uses an EM algorithm to calculate ML estimates of haplotype frequencies given genotype measurements which do not specify phase. The algorithm also allows for some genotype measurements to be missing (due, for example, to PCR failure). It also allows multiple imputation of individual haplotypes (see, e.g., Clayton, D. (2002), "SNPHAP: A Program for Estimating Frequencies of Large Haplotypes of SNPs", which is hereby incorporated by reference in its entirety). An exemplary program is SNPHAP (available at the world wide web at gene.cimr.cam.ac.uk/clayton/software/snphap.txt, which is hereby incorporated by reference in its entirety).

# US 11,530,454 B2

119

In one embodiment, an individual's genetic data is phased using an algorithm that estimates haplotypes from population genotype data, such as an algorithm for haplotype inference based on genotype statistics collected for pairs of SNPs. This software can be used for comparatively accurate 5 phasing of large number of long genome sequences, e.g. obtained from DNA arrays. An exemplary program takes genotype matrix as an input, and outputs the corresponding haplotype matrix (see, e.g., Brinza and Zelikovsky, "2SNP: scalable phasing based on 2-SNP haplotypes," Bioinformatics. 22(3):371-3, 2006, which is hereby incorporated by reference in its entirety). An exemplary program is 2SNP (available at the world wide web at alla.cs.gsu.edu/~software/2SNP, which is hereby incorporated by reference in its entirety).

In various embodiments, an individual's genetic data is phased using data about the probability of chromosomes crossing over at different locations in a chromosome or chromosome segment (such as using recombination data such as may be found in the HapMap database to create a 20 recombination risk score for any interval) to model dependence between polymorphic alleles on the chromosome or chromosome segment. In some embodiments, allele counts at the polymorphic loci are calculated on a computer based on sequencing data or SNP array data. In some embodi- 25 ments, a plurality of hypotheses each pertaining to a different possible state of the chromosome or chromosome segment (such as an overrepresentation of the number of copies of a first homologous chromosome segment as compared to a second homologous chromosome segment in the genome 30 of one or more cells from an individual, a duplication of the first homologous chromosome segment, a deletion of the second homologous chromosome segment, or an equal representation of the first and second homologous chromosome segments) are created (such as creation on a com- 35 puter); a model (such as a joint distribution model) for the expected allele counts at the polymorphic loci on the chromosome is built (such as building on a computer) for each hypothesis; a relative probability of each of the hypotheses is determined (such as determination on a computer) using 40 the joint distribution model and the allele counts; and the hypothesis with the greatest probability is selected. In some embodiments, building a joint distribution model for allele counts and the step of determining the relative probability of each hypothesis are done using a method that does not 45 require the use of a reference chromosome.

In one embodiment, genetic data of an individual is phased using genetic data of one or more relatives of the individual (such as one or more parents, siblings, children, fetuses, embryos, grandparents, uncles, aunts, or cousins). In 50 one embodiment, genetic data of an individual is phased using genetic data of one or more genetic offspring of the individual (e.g., 1, 2, 3, or more offspring), such as embryos, fetuses, born children, or a sample of a miscarriage. In one embodiment, genetic data of a parent (such as a parent of a 55 gestating fetus or embryo) is phased using phased haplotypic data for the other parent along with unphased genetic data of one or more genetic offspring of the parents.

In some embodiments, a sample (e.g., a biopsy such as a tumor biopsy, blood sample, plasma sample, serum sample, 60 or another sample likely to contain mostly or only cells, DNA, or RNA with a CNV of interest) from the individual (such as an individual suspected of having cancer, a fetus, or an embryo) is analyzed to determine the phase for one or more regions that are known or suspected to contain a CNV 65 of interest (such as a deletion or duplication). In some embodiments, the sample has a high tumor fraction (such as

120

30, 40, 50, 60, 70, 80, 90, 95, 98, 99, or 100%). In some embodiments, a sample (e.g., a maternal whole blood sample, cells isolated from a maternal blood sample, maternal plasma sample, maternal serum sample, amniocentesis sample, placental tissue sample (e.g., chorionic villus, decidua, or placental membrane), cervical mucus sample, fetal tissue after fetal demise, other sample from a fetus, or another sample likely to contain mostly or only cells, DNA, or RNA with a CNV of interest) from a fetus or the pregnant mother of a fetus is analyzed to determine the phase for one or more regions that are known or suspected to contain a CNV of interest (such as a deletion or duplication). In some embodiments, the sample has a high fetal fraction (such as 25, 30, 40, 50, 60, 70, 80, 90, 95, 98, 99, or 100%).

In some embodiments, the sample has a haplotypic imbalance or any aneuploidy. In some embodiments, the sample includes any mixture of two types of DNA where the two types have different ratios of the two haplotypes, and share at least one haplotype. For example, in the fetal-maternal case, the mother is 1:1 and the fetus is 1:0 (plus a paternal haplotype). For example, in the tumor case, the normal tissue is 1:1, and the tumor tissue is 1:0 or 1:2, 1:3, 1:4, etc. In some embodiments, at least 10; 100; 500; 1,000; 2,000; 3,000; 5,000; 8,000; or 10,000 polymorphic loci are analyzed to determine the phase of alleles at some or all of the loci. In some embodiments, a sample is from a cell or tissue that was treated to become aneuploidy, such as aneuploidy induced by prolonged cell culture.

In some embodiments, a large percent or all of the DNA or RNA in the sample has the CNV of interest. In some embodiments, the ratio of DNA or RNA from the one or more target cells that contain the CNV of interest to the total DNA or RNA in the sample is at least 80, 85, 90, 95, or 100%. For samples with a deletion, only one haplotype is present for the cells (or DNA or RNA) with the deletion. This first haplotype can be determined using standard methods to determine the identity of alleles present in the region of the deletion. In samples that only contain cells (or DNA or RNA) with the deletion, there will only be signal from the first haplotype that is present in those cells. In samples that also contain a small amount of cells (or DNA or RNA) without the deletion (such as a small amount of noncancerous cells), the weak signal from the second haplotype in these cells (or DNA or RNA) can be ignored. The second haplotype that is present in other cells, DNA, or RNA from the individual that lack the deletion can be determined by inference. For example, if the genotype of cells from the individual without the deletion is (AB,AB) and the phased data for the individual indicates that the first haplotype is (A,A); then, the other haplotype can be inferred to be (B,B).

For samples in which both cells (or DNA or RNA) with a deletion and cells (or DNA or RNA) without a deletion are present, the phase can still be determined. For example, plots can be generated similar to FIG. 18 or 29 in which the x-axis represents the linear position of the individual loci along the chromosome, and the y-axis represents the number of A allele reads as a fraction of the total (A+B) allele reads. In some embodiments for a deletion, the pattern includes two central bands that represent SNPs for which the individual is heterozygous (top band represents AB from cells without the deletion and A from cells with the deletion and B from cells with the deletion and B from cells with the deletion and B from cells with the deletion of these two bands increases as the fraction of cells, DNA, or RNA with the deletion increases. Thus, the identity

## US 11,530,454 B2

12

of the A alleles can be used to determine the first haplotype, and the identity of the B alleles can be used to determine the second haplotype.

For samples with a duplication, an extra copy of the haplotype is present for the cells (or DNA or RNA) with 5 duplication. This haplotype of the duplicated region can be determined using standard methods to determine the identity of alleles present at an increased amount in the region of the duplication, or the haplotype of the region that is not duplicated can be determined using standard methods to 10 determine the identity of alleles present at an decreased amount. Once one haplotype is determined, the other haplotype can be determined by inference.

For samples in which both cells (or DNA or RNA) with a duplication and cells (or DNA or RNA) without a dupli- 15 cation are present, the phase can still be determined using a method similar to that described above for deletions. For example, plots can be generated similar to FIG. 18 or 29 in which the x-axis represents the linear position of the individual loci along the chromosome, and the y-axis represents 20 the number of A allele reads as a fraction of the total (A+B) allele reads. In some embodiments for a deletion, the pattern includes two central bands that represent SNPs for which the individual is heterozygous (top band represents AB from cells without the duplication and AAB from cells with the 25 duplication, and bottom band represents AB from cells without the duplication and ABB from cells with the duplication). In some embodiments, the separation of these two bands increases as the fraction of cells, DNA, or RNA with the duplication increases. Thus, the identity of the A alleles 30 can be used to determine the first haplotype, and the identity of the B alleles can be used to determine the second haplotype. In some embodiments, the phase of one or more CNV region(s) (such as the phase of at least 50, 60, 70, 80, 90, 95, or 100% of the polymorphic loci in the region that 3: were measured) is determined for a sample (such as a tumor biopsy or plasma sample) from an individual known to have cancer and is used for analysis of subsequent samples from the same individual to monitor the progression of the cancer (such as monitoring for remission or reoccurrence of the 40 cancer). In some embodiments, a sample with a high tumor fraction (such as a tumor biopsy or a plasma sample from an individual with a high tumor load) is used to obtain phased data that is used for analysis of subsequent samples with a lower tumor fraction (such as a plasma sample from an 45 individual undergoing treatment for cancer or in remission).

In another embodiment for prenatal diagnostics, phased parental haplotypic data is to detect the presence of more than one homolog from the father, implying that the genetic material from more than one fetus is present in a maternal 50 blood sample. By focusing on chromosomes that are expected to be euploid in a fetus, one could rule out the possibility that the fetus was afflicted with a trisomy. Also, it is possible to determine if the fetal DNA is not from the current father.

In some embodiments, two or more of the methods described herein are used to phase genetic data of an individual. In some embodiments, both a bioinformatics method (such as using population based haplotype frequencies to infer the most likely phase) and a molecular biology 60 method (such as any of the molecular phasing methods disclosed herein to obtain actual phased data rather than bioinformatics-based inferred phased data) are used. In some embodiments, phased data from other subjects (such as prior subjects) is used to refine the population data. For 65 example, phased data from other subjects can be added to population data to calculate priors for possible haplotypes

122

for another subject. In some embodiments, phased data from other subjects (such as prior subjects) is used to calculate priors for possible haplotypes for another subject.

In some embodiments, probabilistic data may be used. For example, due to the probabilistic nature of the representation of DNA molecules in a sample, as well as various amplification and measurement biases, the relative number of molecules of DNA measured from two different loci, or from different alleles at a given locus, is not always representative of the relative number of molecules in the mixture, or in the individual. If one were trying to determine the genotype of a normal diploid individual at a given locus on an autosomal chromosome by sequencing DNA from the plasma of the individual, one would expect to either observe only one allele (homozygous) or about equal numbers of two alleles (heterozygous). If, at that allele, ten molecules of the A allele were observed, and two molecules of the B allele were observed, it would not be clear if the individual was homozygous at the locus, and the two molecules of the B allele were due to noise or contamination, or if the individual was heterozygous, and the lower number of molecules of the B allele were due to random, statistical variation in the number of molecules of DNA in the plasma, amplification bias, contamination or any number of other causes. In this case, a probability that the individual was homozygous, and a corresponding probability that the individual was heterozygous could be calculated, and these probabilistic genotypes could be used in further calculations.

Note that for a given allele ratio, the likelihood that the ratio closely represents the ratio of the DNA molecules in the individual is greater the greater the number of molecules that are observed. For example, if one were to measure 100 molecules of A and 100 molecules of B, the likelihood that the actual ratio was 50% is considerably greater than if one were to measure 10 molecules of A and 10 molecules of B. In one embodiment, one uses Bayesian theory combined with a detailed model of the data to determine the likelihood that a particular hypothesis is correct given an observation. For example, if one were considering two hypotheses—one that corresponds to a trisomic individual and one that corresponds to a disomic individual—then the probability of the disomic hypothesis being correct would be considerably higher for the case where 100 molecules of each of the two alleles were observed, as compared to the case where 10 molecules of each of the two alleles were observed. As the data becomes noisier due to bias, contamination or some other source of noise, or as the number of observations at a given locus goes down, the probability of the maximum likelihood hypothesis being true given the observed data drops. In practice, it is possible to aggregate probabilities over many loci to increase the confidence with which the maximum likelihood hypothesis may be determined to be the correct hypothesis. In some embodiments, the probabilities are simply aggregated without regard for recombination. In some embodiments, the calculations take into account cross-overs.

In an embodiment, probabilistically phased data is used in the determination of copy number variation. In some embodiments, the probabilistically phased data is population based haplotype block frequency data from a data source such as the HapMap data base. In some embodiments, the probabilistically phased data is haplotypic data obtained by a molecular method, for example phasing by dilution where individual segments of chromosomes are diluted to a single molecule per reaction, but where, due to stochastic noise the identities of the haplotypes may not be absolutely known. In some embodiments, the probabilistically phased data is

## US 11,530,454 B2

123

haplotypic data obtained by a molecular method, where the identities of the haplotypes may be known with a high degree of certainty.

Imagine a hypothetical case where a doctor wanted to determine whether or not an individual had some cells in 5 their body which had a deletion at a particular chromosomal segment by measuring the plasma DNA from the individual. The doctor could make use of the knowledge that if all of the cells from which the plasma DNA originated were diploid, and of the same genotype, then for heterozygous loci, the relative number of molecules of DNA observed for each of the two alleles would fall into one distribution that was centered at 50% A allele and 50% B allele. However, if a fraction of the cells from which the plasma DNA originated had a deletion at a particular chromosome segment, then for heterozygous loci, one would expect that the relative number of molecules of DNA observed for each of the two alleles would fall into two distributions, one centered at above 50% A allele for the loci where there was a deletion 20 of the chromosome segment containing the B allele, and one centered at below 50% for the loci where there was a deletion of the chromosome segment containing the A allele. The greater the proportion of the cells from which the plasma DNA originated contained the deletion, the further 25 from 50% these two distributions would be.

In this hypothetical case, imagine a clinician who wants to determine if an individual had a deletion of a chromosomal region in a proportion of cells in the individual's body. The clinician may draw blood from the individual into 30 a vacutainer or other type of blood tube, centrifuge the blood, and isolate the plasma layer. The clinician may isolate the DNA from the plasma, enrich the DNA at the targeted loci, possibly through targeted or other amplification, locus capture techniques, size enrichment, or other enrichment techniques. The clinician may analyze such as by measuring the number of alleles at a set of SNPs, in other words generating allele frequency data, the enriched and/or amplified DNA using an assay such as qPCR, sequencing, a microarray, or other techniques that measure the quantity of 40 DNA in a sample. We will consider data analysis for the case where the clinician amplified the cell-free plasma DNA using a targeted amplification technique, and then sequenced the amplified DNA to give the following exemplary possible data at six SNPs found on a chromosome segment that is 45 indicative of cancer, where the individual was heterozygotic at those SNPs:

SNP 1: 460 reads A allele; 540 reads B allele (46% A) SNP 2: 530 reads A allele; 470 reads B allele (53% A) SNP 3: 40 reads A allele; 60 reads B allele (40% A) SNP 4: 46 reads A allele; 54 reads B allele (46% A) SNP 5: 520 reads A allele; 480 reads B allele (52% A) SNP 6: 200 reads A allele; 200 reads B allele (50% A)

From this set of data, it may be difficult to differentiate between the case where the individual is normal, with all 55 cells being disomic, or where the individual may have a cancer, with some portion of cells whose DNA contributed towards the cell-free DNA found in the plasma having a deletion or duplication at the chromosome. For example, the two hypotheses with the maximum likelihood may be that 60 the individual has a deletion at this chromosome segment, with a tumor fraction of 6%, and where the deleted segment of the chromosome has the genotype over the six SNPs of (A,B,A,A,B,B) or (A,B,A,A,B,A). In this representation of the individual's genotype over a set of SNPs, the first letter in the parentheses corresponds to the genotype of the haplotype for SNP 1, the second to SNP 2, etc.

124

If one were to use a method to determine the haplotype of the individual at that chromosome segment, and were to find that the haplotype for one of the two chromosomes was (A,B,A,A,B,B), this would agree with the maximum likelihood hypothesis, and the calculated likelihood that the individual has a deletion at that segment, and therefore may have cancerous or precancerous cells, would be considerably increased. On the other hand, if the individual were found to have the haplotype (A,A,A,A,A), then the likelihood that the individual has a deletion at that chromosome segment would be considerably decreased, and perhaps the likelihood of the no-deletion hypothesis would be higher (the actual likelihood values would depend on other parameters such as the measured noise in the system, among others).

There are many ways to determine the haplotype of the individual, many of which are described elsewhere in this document. A partial list is given here, and is not meant to be exhaustive. One method is a biological method where individual DNA molecules are diluted until approximately one molecule from each chromosomal region is in any given reaction volume, and then methods such as sequencing are used to measure the genotype. Another method is informatically based where population data on various haplotypes coupled with their frequency can be used in a probabilistic manner. Another method is to measure the diploid data of the individual, along with one or a plurality of related individuals who are expected to share haplotype blocks with the individual and to infer the haplotype blocks. Another method would be to take a sample of tissue with a high concentration of the deleted or duplicated segment, and determine the haplotype based on allelic imbalance, for example, genotype measurements from a sample of tumor tissue with a deletion can be used to determine the phased data for that deletion region, and this data can then be used to determine if the cancer has regrown post-resection.

In practice, typically more than 20 SNPs, more than 50 SNPs, more than 100 SNPs, more than 500 SNPs, more than 1,000 SNPs, or more than 5,000 SNPs are measured on a given chromosome segment.

Exemplary Methods for Phasing, Predicting Allele Ratios, and Reconstructing Fetal Genetic Data

In one aspect, the invention features methods for determining one or more haplotypes of a fetus. In various embodiments, this method allows one to determine which polymorphic loci (such as SNPs) were inherited by the fetus and to reconstruct which homologs (including recombination events) are present in the fetus (and thereby interpolate the sequence between the polymorphic loci). If desired, essentially the entire genome of the fetus can be reconstructed. If there is some remaining ambiguity in the genome of the fetus (such as in intervals with a crossover), this ambiguity can be minimized if desired by analyzing additional polymorphic loci. In various embodiments, the polymorphic loci are chosen to cover one or more of the chromosomes at a density to reduce any ambiguity to a desired level. This method has important applications for the detection of polymorphisms or other mutations of interest (such as deletions or duplications) in a fetus since it enables their detection based on linkage (such as the presence of linked polymorphic loci in the fetal genome) rather than by directing detecting the polymorphism or other mutation of interest in the fetal genome. For example, if a parent is a carrier for a mutation associated with cystic fibrosis (CF), a nucleic acid sample that includes maternal DNA from the mother of the fetus and fetal DNA from the fetus can be analyzed to determine whether the fetal DNA include the Case: 24-1324 Document: 42-1 Page: 512 Filed: 03/18/2024

## US 11,530,454 B2

125

haplotype containing the CF mutation. In particular, polymorphic loci can be analyzed to determine whether the fetal DNA includes the haplotype containing the CF mutation without having to detect the CF mutation itself in the fetal DNA. This is useful in screening for one or more mutations, such as disease-linked mutations, without having to directly detect the mutations.

In some embodiments, the method involves determining a parental haplotype (e.g., a haplotype of the mother or father of the fetus), such as by using any of the methods described herein. In some embodiments, this determination is made without using data from a relative of the mother or father. In some embodiments, a parental haplotype is determined using a dilution approach followed by SNP genotyping or sequencing as described herein. In some embodiments, a haplotype of the mother (or father) is determined by any of the methods described herein using data from a relative of the mother (or father). In some embodiments, a haplotype is determined for both the father and the mother. 20

This parental haplotype data can be used to determine if the fetus inherited the parental haplotype. In some embodiments, a nucleic acid sample that includes maternal DNA from the mother of the fetus and fetal DNA from the fetus 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci. In some embodiments, a nucleic acid sample that includes maternal DNA from the mother of the fetus and fetal DNA from the fetus is analyzed by contacting the sample with a library of primers that simultaneously hybridize to at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci (such as SNPs) to produce a reaction mixture. In some embodiments, the reaction mixture is subjected to primer extension reaction conditions to produce amplified products. In some embodiments, the amplified products are measured with a high throughput sequencer to produce sequencing data.

In various embodiments, a fetal haplotype is determined using data about the probability of chromosomes crossing over at different locations in a chromosome or chromosome segment (such as by using recombination data such as may be found in the HapMap database to create a recombination 45 risk score for any interval) to model dependence between polymorphic alleles on the chromosome or chromosome segment as described above. In some embodiments, the method takes into account physical distance of the SNPs (such as SNPs flanking a gene or mutation of interest) and 50 recombination data from location specific recombination likelihoods and the data observed from the genetic measurements of the maternal plasma to obtain the most likely genotype of the fetus. Then PARENTAL SUPPORT<sup>TM</sup> may be performed on the targeted sequencing or SPN array data 55 obtained from these SNPs to determine which homologs were inherited by the fetus from both parents (see, e.g., U.S. application Ser. No. 11/603,406 (US Publication No. 20070184467), U.S. application Ser. No. 12/076,348 (US Publication No. 20080243398), U.S. application Ser. No. 60 13/110,685 (U.S. Publication No. 2011/0288780), PCT Application PCT/US09/52730 (PCT Publication No. WO/2010/017214), and PCT Application No. PCT/US10/ 050824 (PCT Publication No. WO/2011/041485), U.S. application Ser. No. 13/300,235 (U.S. Publication No. 2012/ 65 0270212), U.S. application Ser. No. 13/335,043 (U.S. Publication No. 2012/0122701), U.S. application Ser. No.

13/683,604, and U.S. application Ser. No. 13/780,022, which are each hereby incorporated by reference in its entirety).

Assume a generalized example where the possible alleles at one locus are A and B; assignment of the identity A or B to particular alleles is arbitrary. Parental genotypes for a particular SNP, termed genetic contexts, are expressed as maternal genotype paternal genotype. Thus, if the mother is homozygous and the father is heterozygous, this would be represented as AAIAB. Similarly, if both parents are homozygous for the same allele, the parental genotypes would be represented as AA|AA. Furthermore, the fetus would never have AB or BB states and the number of sequence reads with the B allele will be low, and thus can be used to determine the noise responses of the assay and genotyping platform, including effects such as low level DNA contamination and sequencing errors; these noise responses are useful for modeling expected genetic data profiles. There are only five possible maternal|paternal genetic contexts: AA|AA, AA|AB, AB|AA, AB|AB, and AA|BB; other contexts are equivalent by symmetry. SNPs where the parents are homozygous for the same allele are only informative for determining noise and contamination levels. SNPs where the parents are not homozygous for the is analyzed using a SNP array to detect at least 100; 200; 25 same allele are informative in determining fetal fraction and copy number count.

Let  $N_{A,i}$  and  $N_{B,i}$  represent the number of reads of each allele at SNP i, and let Ci represent the parental genetic context at that locus. The data set for a particular chromosome is represented by  $N_{AB} = \{N_{A,i}, N_{B,i}\}\ i=1 \dots N$  and  $C=\{C_i\}$ , i=1 ... N. For reconstructing part or all of the fetal genome, it can optionally be determined if the fetus has an aneuploidy (such as a missing or extra copy of a chromosome or chromosome segment). For each individual chromosome or chromosome under study, let H represent the set of one or more hypotheses for the total number of chromosomes, the parental origin of each chromosome, and the positions on the parent chromosomes where recombination occurred during formation of the gametes that fertilized to create the child. The probability of a hypothesis P(H) can be computed using the data from the HapMap database and prior information related to each of the ploidy states.

Furthermore, let F represent the fetal cfDNA fraction in the sample. Given a set of possible H, C, and F, one can compute the probability of  $N_{AB}$ ,  $P(N_{AB}|H,F,C)$  based on modeling the noise sources of the molecular assay and the sequencing platform. The goal is to find the hypothesis H and the fetal fraction F that maximizes  $P(H,F|N_{AB})$ . Using standard Bayesian statistical techniques, and assuming a uniform probability distribution for F from 0 to 1, this can be recast in terms of maximizing the probability of  $P(N_{AB}|H,$ F,C)P(H) over H and F, all of which can now be computed. The probability of all hypotheses associated with a particular copy number and fetal fraction, e.g., trisomy and F=10%, but covering all possible parental chromosome origins and crossover locations, are summed. The copy number hypothesis with the highest probability is selected as the test result, the fetal fraction associated with that hypothesis reveals the fetal fraction, and the probability associated with that hypothesis is the calculated accuracy of the result.

In some embodiments, the algorithm uses in silico simulations to generate a very large number of hypothetical sequencing data sets that could result from the possible fetal genetic inheritance patterns, sample parameters, and amplification and measurement artifacts of the method. More specifically, the algorithm first utilizes parental genotypes at a large number of SNPs and crossover frequency data from

## US 11,530,454 B2

127

the HapMap database to predict possible fetal genotypes. It then predicts expected data profiles for the sequencing data that would be measured from mixed samples originating from a mother carrying a fetus with each of the possible fetal genotypes and taking into account a variety of parameters including fetal fraction, expected read depth profile, fetal genome equivalents present in the sample, expected amplification bias at each of the SNPs, and a number of noise parameters. A data model describes how the sequencing or SNP array data is expected to appear for each of these phypotheses given the particular parameter set. The hypothesis with the best data fit between this modeled data and the measured data is selected.

If desired, expected allele ratios can be calculated for DNA or RNA from the fetus using the results of what haplotypes were inherited by the fetus. The expected allele ratios can also be calculated for a mixed sample containing nucleic acids from both the mother and the fetus (these allele ratios indicate what is expected for measurement of the total amount of each allele, including the amount of the allele from both maternal nucleic acids and fetal nucleic acids in the sample). The expected allele ratios can be calculated for different hypotheses specifying the degree of overrepresentation of the first homologous chromosome segment.

In some embodiments, the method involves determining 2 whether the fetus has one or more of the following conditions: cystic fibrosis, Huntington's disease, Fragile X, thalassemia, muscular dystrophy (such as Duchenne's muscular dystrophy), Alzheimer, Fanconi Anemia, Gaucher Disease, Mucolipidosis IV, Niemann-Pick Disease, Tay-Sachs dis- 3 ease, Sickle cell anemia, Parkinson disease, Torsion Dystonia, and cancer. In some embodiments, a fetal haplotype is determined for one or more chromosomes taken from the group consisting of chromosomes 13, 18, 21, X, and Y. In some embodiments, a fetal haplotype is determined for all of the fetal chromosomes. In various embodiments, the method determines essentially the entire genome of the fetus. In some embodiments, the haplotype is determined for at least 30, 40, 50, 60, 70, 80, 90, or 95% of the genome of the fetus. In some embodiments, the haplotype determination of the 4 fetus includes information about which allele is present for at least 100; 200; 500; 750; 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different polymorphic loci. In some embodiments, this method is used to determine a haplotype or allele ratios 4 for an embryo.

Exemplary Methods for Predicting Allele Ratios

Exemplary methods are described below for calculating expected allele ratios for a sample. Table 1 shows expected allele ratios for a mixed sample (such as a maternal blood 5 sample) containing nucleic acids from both the mother and the fetus. These expected allele ratios indicate what is expected for measurement of the total amount of each allele, including the amount of the allele from both maternal nucleic acids and fetal nucleic acids in the mixed sample. In 5 an example, the mother is heterozygous at two neighboring loci that are expected to cosegregate (e.g., two loci for which no chromosome crossovers are expected between the loci). Thus, the mother is (AB, AB). Now imagine that the phased data for the mother indicates that for one haplotype she is (A, 6 A); thus, for the other haplotype one can infer that she is (B, B). Table 1 gives the expected allele ratios for different hypotheses where the fetal fraction is 20%. For this example, no knowledge of the paternal data is assumed, and the heterozygosity rate is assumed to be 50%. The expected 6 allele ratios are given in terms of (expected proportion of A reads/total number of reads) for each of the two SNPs. These

128

ratios are calculated both using maternal phased data (the knowledge that one haplotype is (A, A) and one is (B, B)) and without using the maternal phased data. Table 1 includes different hypotheses for the number of copies of the chromosome segment in the fetus from each parent.

TABLE 1

Expected Genetic Data for Mixed Sample of Maternal and Fetal Nucleic Acids		
Copy Number Hypothesis	Expected allele ratios when using maternal phased data	Expected allele ratios when not usin maternal phased data
Monosomy	(0.444; 0.444)	(0.444; 0.444)
(maternal copy missing)	(0.444; 0.555)	(0.444; 0.555)
	(0.555; 0.444)	(0.555; 0.444)
	(0.555; 0.555)	(0.555; 0.555)
Monosomy	(0.444; 0.444) (0.555; 0.555)	(0.444; 0.444)
(paternal copy missing)	(0.555; 0.555)	(0.444; 0.555) (0.555; 0.444)
		(0.555; 0.555)
Disomy	(0.40; 0.40)	(0.40; 0.40)
	(0.40; 0.50)	(0.40; 0.50)
	(0.50; 0.40)	(0.40; 0.60)
	(0.50; 0.50)	(0.50; 0.40)
	(0.50; 0.60)	(0.50; 0.50)
	(0.60; 0.50)	(0.50; 0.60)
	(0.60; 0.60)	(0.60; 0.40)
		(0.60; 0.50) (0.60; 0.60)
Гrisomy	(0.36; 0.36)	(0.36; 0.36)
extra matching	(0.36; 0.45)	(0.36; 0.45)
maternal copy)	(0.45; 0.36)	(0.36; 0.54)
	(0.45; 0.45)	(0.36; 0.63)
	(0.54; 0.54)	(0.45; 0.36)
	(0.54; 0.63)	(0.45; 0.45)
	(0.63; 0.54)	(0.45; 0.54)
	(0.63; 0.63)	(0.45; 0.63)
		(0.54; 0.36) (0.54; 0.45)
		(0.54; 0.54)
		(0.54; 0.63)
		(0.63; 0.36)
		(0.63; 0.45)
Trisomy (extra unmatching		(0.63; 0.54)
	/ ·-	(0.63; 0.63)
	(0.45, 0.45)	(0.36; 0.36)
	(0.45; 0.54) (0.54; 0.45)	(0.36; 0.45) (0.36; 0.54)
maternal copy)	(0.54; 0.54)	(0.36; 0.63)
	(0.54, 0.54)	(0.45; 0.36)
		(0.45; 0.45)
		(0.45; 0.54)
		(0.45; 0.63)
		(0.54; 0.36)
		(0.54; 0.45)
Trisomy		(0.54; 0.54)
		(0.54; 0.63) (0.63; 0.36)
		(0.63; 0.45)
		(0.63; 0.54)
		(0.63; 0.63)
	(0.36; 0.36)	(0.36; 0.36)
extra matching	(0.36; 0.54)	(0.36; 0.45)
paternal copy)	(0.54; 0.36)	(0.36; 0.54)
	(0.54; 0.54)	(0.36; 0.63)
	(0.45; 0.45)	(0.45; 0.36)
	(0.45; 0.63)	(0.45; 0.45)
	(0.63; 0.45) (0.63; 0.63)	(0.45; 0.54) (0.45; 0.63)
	(0.05, 0.05)	(0.43; 0.63)
		(0.54; 0.45)
		(0.54; 0.54)
		(0.54; 0.63)
		(0.63; 0.36)
		(0.63; 0.45) (0.63; 0.54)

## US 11,530,454 B2

129
TABLE 1-continued

Expected Genetic Data for Mixed Sample of Maternal and Fetal Nucleic Acids		
Copy Number Hypothesis	Expected allele ratios when using maternal phased data	Expected allele ratios when not using maternal phased data
Trisomy (extra unmatching paternal copy)	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.36) (0.45; 0.54) (0.45; 0.54) (0.54; 0.36) (0.54; 0.36) (0.54; 0.45) (0.54; 0.54) (0.54; 0.63) (0.54; 0.63) (0.63; 0.36) (0.63; 0.36) (0.63; 0.45) (0.63; 0.54) (0.63; 0.54)	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.36) (0.45; 0.54) (0.45; 0.63) (0.54; 0.36) (0.54; 0.36) (0.54; 0.54) (0.54; 0.63) (0.54; 0.63) (0.54; 0.63) (0.63; 0.63)

In addition to the fact that using phased data reduces the number of possible expected allele ratios, it also changes the prior likelihood of each of the expected allele ratios, such that the maximum likelihood result is more likely to be correct. Eliminating expected allele ratios or hypotheses that are not possible increases the likelihood that the correct hypothesis will be chosen. As an example, suppose the measured allele ratios are (0.41, 0.59). Without using phased data, one might assume that the hypothesis with maximum likelihood is a disomy hypothesis (given the similarity of the measured allele ratios to expected allele ratios of (0.40, 0.60) for disomy). However, using phased data, one can exclude (0.40, 0.60) as expected allele ratios for the disomy hypothesis, and one can select a trisomy hypothesis as more likely.

Assume the measured allele ratios are (0.4, 0.4). Without any haplotype information, the probability of a maternal deletion at each SNP would be the 0.5×P(A deleted)+0.5× P(B deleted). Therefore, although it looks like A is deleted (missing in the fetus), the likelihood of deletion would be the average of the two. For high enough fetal fraction, one can still determine the most likely hypothesis. For low enough fetal fraction, averaging may work in disfavor of the deletion hypothesis. However, with haplotype information, the probability of homolog 1 being deleted, P (A deleted), is greater and will fit the measured data better. If desired, crossover probabilities between the two loci can also be considered.

In a further illustrative example of combining likelihoods using phased data, consider two consecutive SNPs s1 and s2, and D1 and D2 denote the allele data in these SNPs. Here we provide an example on how to combine the likelihoods for these two SNPs. Let c denote the probability that two consecutive heterozygous SNPs have the same allele in the same homolog (i.e., both SNPs are AB or both SNPs are BA). Hence 1-c is the probability that one SNP is AB and the other one is BA. For example, consider the hypothesis H10 and allelic imbalance value f. First, assume that all likelihoods are computed assuming that all SNPs are either AB or BA. Then, we can combine the likelihoods in two consecutive SNPs as follows:

$$\begin{split} \operatorname{Lik}(D_{1}, & D_{2} | H_{10}, f) = \operatorname{Lik}(D_{1} | H_{10}, f) \times c \times \operatorname{Lik}(D_{2} | H_{10}, f) + \operatorname{Lik} \\ & (D_{1} | H_{10}, f) \times (1 - c) \times \operatorname{Lik}(D_{2} | H_{01}, f). \end{split}$$

We can do this recursively to determine the joint likelihood  $Lik(D_1, \ldots, D_N|H_{10}, f)$  for all SNPs.

#### 130

**Exemplary Mutations** 

Exemplary mutations associated with a disease or disorder such as cancer or an increased risk (such as an above normal level of risk) for a disease or disorder such as cancer include single nucleotide variants (SNVs), multiple nucleotide mutations, deletions (such as deletion of a 2 to 30 million base pair region), duplications, or tandem repeats. In some embodiments, the mutation is in DNA, such as cfDNA, cell-free mitochondrial DNA (cf mDNA), cell-free DNA that originated from nuclear DNA (cf nDNA), cellular DNA, or mitochondrial DNA. In some embodiments, the mutation is in RNA, such as cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA. In some embodiments, the mutation is present at <sup>15</sup> a higher frequency in subjects with a disease or disorder (such as cancer) than subjects without the disease or disorder (such as cancer). In some embodiments, the mutation is indicative of cancer, such as a causative mutation. In some embodiments, the mutation is a driver mutation that has a 20 causative role in the disease or disorder. In some embodiments, the mutation is not a causative mutation. For example, in some cancers, multiple mutations accumulate but some of them are not causative mutations. Mutations (such as those that are present at a higher frequency in subjects with a disease or disorder than subjects without the disease or disorder) that are not causative can still be useful for diagnosing the disease or disorder. In some embodiments, the mutation is loss-of-heterozygosity (LOH) at one or more microsatellites.

In some embodiments, a subject is screened for one of more polymorphisms or mutations that the subject is known to have (e.g., to test for their presence, a change in the amount of cells, DNA, or RNA with these polymorphisms or mutations, or cancer remission or re-occurrence). In some embodiments, a subject is screened for one of more polymorphisms or mutations that the subject is known to be at risk for (such as a subject who has a relative with the polymorphism or mutation). In some embodiments, a subject is screened for a panel of polymorphisms or mutations associated with a disease or disorder such as cancer (e.g., at least 5, 10, 50, 100, 200, 300, 500, 750, 1,000, 1,500, 2,000, or 5,000 polymorphisms or mutations).

Many coding variants associated with cancer are described in Abaan et al., "The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology", Cancer Research, Jul. 15, 2013, and world wide web at dtp.nci.nih.gov/branches/btb/characterizationNCI60.html, which are each hereby incorporated by reference in its entirety). The NCI-60 human cancer cell line panel consists of 60 different cell lines representing cancers of the lung, colon, brain, ovary, breast, prostate, and kidney, as well as leukemia and melanoma. The genetic variations that were identified in these cell lines consisted of two types: type I variants that are found in the normal population, and type II variants that are cancer-specific.

Exemplary polymorphisms or mutations (such as deletions or duplications) are in one or more of the following genes: TP53, PTEN, PIK3CA, APC, EGFR, NRAS, NF2, FBXW7, ERBBs, ATAD5, KRAS, BRAF, VEGF, EGFR, 60 HER2, ALK, p53, BRCA, BRCA1, BRCA2, SETD2, LRP1B, PBRM, SPTA1, DNMT3A, ARID1A, GRIN2A, TRRAP, STAG2, EPHA3/5/7, POLE, SYNE1, C20orf80, CSMD1, CTNNB1, ERBB2. FBXW7, KIT, MUC4, ATM, CDH1, DDX11, DDX12, DSPP, EPPK1, FAM186A, 65 GNAS, HRNR, KRTAP4-11, MAP2K4, MLL3, NRAS, RB1, SMAD4, TTN, ABCC9, ACVR1B, ADAM29, ADAMTS19, AGAP10, AKT1, AMBN, AMPD2,

## US 11,530,454 B2

13

ANKRD30A, ANKRD40, APOBR, AR, BIRC6, BMP2, BRAT1, BTNL8, C12orf4, C1QTNF7, C20orf186, CAP-RIN2, CBWD1, CCDC30, CCDC93, CD5L, CDC27, CDC42BPA, CDH9, CDKN2A, CHD8, CHEK2, CHRNA9, CIZ1, CLSPN, CNTN6, COL14A1, CREBBP, CROCC, CTSF, CYP1A2, DCLK1, DHDDS, DHX32, DKK2, DLEC1, DNAH14, DNAH5, DNAH9, DNASE1L3, DUSP16, DYNC2H1, ECT2, EFHB, RRN3P2, TRIM49B, TUBB8P5, EPHA7, ERBB3, ERCC6, FAM21A, FAM21C, FCGBP, FGFR2, FLG2, FLT1, FOLR2, FRYL, FSCB, 10 GAB1, GABRA4, GABRP, GH2, GOLGA6L1, GPHB5, GPR32, GPX5, GTF3C3, HECW1, HIST1H3B, HLA-A, HRAS, HS3ST1, HS6ST1, HSPD1, IDH1, JAK2, KDM5B, KIAA0528, KRT15, KRT38, KRTAP21-1, KRTAP4-5, KRTAP4-7, KRTAP5-4, KRTAP5-5, LAMA4, LATS1, 15 LMF1, LPAR4, LPPR4, LRRFIP1, LUM, LYST, MAP2K1, MARCH1, MARCO, MB21D2, MEGF10, MMP16, MORC1, MRE11A, MTMR3, MUC12, MUC17, MUC2, MUC20, NBPF10, NBPF20, NEK1, NFE2L2, NLRP4, NOTCH2, NRK, NUP93, OBSCN, OR11H1, OR2B11, 20 OR2M4, OR4Q3, OR5D13, OR812, OXSM, PIK3R1, PPP2R5C, PRAME, PRF1, PRG4, PRPF19, PTH2, PTPRC, PTPRJ, RAC1, RAD50, RBM12, RGPD3, RGS22, ROR1, RP11-671M22.1, RP13-996F3.4, RP1L1, RSBN1L, RYR3, SAMD3, SCN3A, SEC31A, SF1, SF3B1, SLC25A2, 25 SLC44A1, SLC4A11, SMAD2, SPTA1, ST6GAL2, STK11, SZT2, TAF1L, TAX1BP1, TBP, TGFBI, TIF1, TMEM14B, TMEM74, TPTE, TRAPPC8, TRPS1, TXNDC6, USP32, UTP20, VASN, VPS72, WASH3P, WWTR1, XPO1, ZFHX4, ZMIZ1, ZNF167, ZNF436, ZNF492, ZNF598, 30 ZRSR2, ABL1, AKT2, AKT3, ARAF, ARFRP1, ARID2, ASXL1, ATR, ATRX, AURKA, AURKB, AXL, BAP1, BARD1, BCL2, BCL2L2, BCL6, BCOR, BCORL1, BLM, BRIP1, BTK, CARD11, CBFB, CBL, CCND1, CCND2, CCND3, CCNE1, CD79A, CD79B, CDC73, CDK12, CDK4, CDK6, CDK8, CDKN1B, CDKN2B, CDKN2C, CEBPA, CHEK1, CIC, CRKL, CRLF2, CSF1R, CTCF, CTNNA1, DAXX, DDR2, DOT1L, EMSY (C11orf30), EP300, EPHA3, EPHA5, EPHB1, ERBB4, ERG, ESR1, EZH2, FAM123B (WTX), FAM46C, FANCA, FANCC, 40 FANCD2, FANCE, FANCF, FANCG, FANCL, FGF10, FGF14, FGF19, FGF23, FGF3, FGF4, FGF6, FGFR1, FGFR2, FGFR3, FGFR4, FLT3, FLT4, FOXL2, GATA1, GATA2, GATA3, GID4 (C17orf39), GNA11, GNA13, GNAQ, GNAS, GPR124, GSK3B, HGF, IDH1, IDH2, 45 IGF1R, IKBKE, IKZF1, IL7R, INHBA, IRF4, IRS2, JAK1, JAK3, JUN, KAT6A (MYST3), KDM5A, KDM5C, KDM6A, KDR, KEAP1, KLHL6, MAP2K2, MAP2K4, MAP3K1, MCL1, MDM2, MDM4, MED12, MEF2B, MEN1, MET, MITF, MLH1, MLL, MLL2, MPL, MSH2, 50 MSH6, MTOR, MUTYH, MYC, MYCL1, MYCN, MYD88, NF1, NFKBIA, NKX2-1, NOTCH1, NPM1, NRAS, NTRK1, NTRK2, NTRK3, PAK3, PALB2, PAX5, PBRM1, PDGFRA, PDGFRB, PDK1, PIK3CG, PIK3R2, PPP2R1A, PRDM1, PRKAR1A, PRKDC, PTCH1, 55 PTPN11, RAD51, RAF1, RARA, RET, RICTOR, RNF43, RPTOR, RUNX1, SMARCA4, SMARCB1, SMO, SOCS1, SOX10, SOX2, SPEN, SPOP, SRC, STAT4, SUFU, TET2, TGFBR2, TNFAIP3, TNFRSF14, TOP1, TP53, TSC1, TSC2, TSHR, VHL, WISP3, WT1, ZNF217, ZNF703, and 60 combinations thereof (Su et al., J Mol Diagn 2011, 13:74-84; DOI:10.1016/j.jmoldx.2010.11.010; and Abaan et al., "The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology", Cancer Research, Jul. 15, 2013, which are each hereby incorporated 65 by reference in its entirety). In some embodiments, the duplication is a chromosome 1p ("Chr1p") duplication asso132

ciated with breast cancer. In some embodiments, one or more polymorphisms or mutations are in BRAF, such as the V600E mutation. In some embodiments, one or more polymorphisms or mutations are in K-ras. In some embodiments, there is a combination of one or more polymorphisms or mutations in K-ras and APC. In some embodiments, there is a combination of one or more polymorphisms or mutations in K-ras and p53. In some embodiments, there is a combination of one or more polymorphisms or mutations in APC and p53. In some embodiments, there is a combination of one or more polymorphisms or mutations in K-ras, APC, and p53. In some embodiments, there is a combination of one or more polymorphisms or mutations in K-ras and EGFR. Exemplary polymorphisms or mutations are in one or more of the following microRNAs: miR-15a, miR-16-1, miR-23a, miR-23b, miR-24-1, miR-24-2, miR-27a, miR-27b, miR-29b-2, miR-29c, miR-146, miR-155, miR-221, miR-222, and miR-223 (Calin et al. "A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia." N Engl J Med 353:1793-801, 2005, which is hereby incorporated by reference in its entirety).

In some embodiments, the deletion is a deletion of at least 0.01 kb, 0.1 kb, 1 kb, 10 kb, 100 kb, 1 mb, 2 mb, 3 mb, 5 mb, 10 mb, 15 mb, 20 mb, 30 mb, or 40 mb. In some embodiments, the deletion is a deletion of between 1 kb to 40 mb, such as between 1 kb to 100 kb, 100 kb to 1 mb, 1 to 5 mb, 5 to 10 mb, 10 to 15 mb, 15 to 20 mb, 20 to 25 mb, 25 to 30 mb, or 30 to 40 mb, inclusive.

In some embodiments, the duplication is a duplication of at least 0.01 kb, 0.1 kb, 1 kb, 10 kb, 100 kb, 1 mb, 2 mb, 3 mb, 5 mb, 10 mb, 15 mb, 20 mb, 30 mb, or 40 mb. In some embodiments, the duplication is a duplication of between 1 kb to 40 mb, such as between 1 kb to 100 kb, 100 kb to 1 mb, 1 to 5 mb, 5 to 10 mb, 10 to 15 mb, 15 to 20 mb, 20 to 25 mb, 25 to 30 mb, or 30 to 40 mb, inclusive.

In some embodiments, the tandem repeat is a repeat of between 2 and 60 nucleotides, such as 2 to 6, 7 to 10, 10 to 20, 20 to 30, 30 to 40, 40 to 50, or 50 to 60 nucleotides, inclusive. In some embodiments, the tandem repeat is a repeat of 2 nucleotides (dinucleotide repeat). In some embodiments, the tandem repeat is a repeat of 3 nucleotides (trinucleotide repeat).

In some embodiments, the polymorphism or mutation is prognostic. Exemplary prognostic mutations include K-ras mutations, such as K-ras mutations that are indicators of post-operative disease recurrence in colorectal cancer (Ryan et al." A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up," Gut 52:101-108, 2003; and Lecomte T et al. Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis," Int J Cancer 100:542-548, 2002, which are each hereby incorporated by reference in its entirety).

In some embodiments, the polymorphism or mutation is associated with altered response to a particular treatment (such as increased or decreased efficacy or side-effects). Examples include K-ras mutations are associated with decreased response to EGFR-based treatments in non-small cell lung cancer (Wang et al. "Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer," Clin Canc Res16: 1324-1330, 2010, which is hereby incorporated by reference in its entirety).

K-ras is an oncogene that is activated in many cancers. Exemplary K-ras mutations are mutations in codons 12, 13, and 61. K-ras cfDNA mutations have been identified in Case: 24-1324 Document: 42-1 Page: 516 Filed: 03/18/2024

## US 11,530,454 B2

133

pancreatic, lung, colorectal, bladder, and gastric cancers (Fleischhacker & Schmidt "Circulating nucleic acids (CNAs) and caner—a survey," Biochim Biophys Acta 1775: 181-232, 2007, which is hereby incorporated by reference in its entirety).

p53 is a tumor suppressor that is mutated in many cancers and contributes to tumor progression (Levine & Oren "The first 30 years of p53: growing ever more complex. Nature Rev Cancer," 9:749-758, 2009, which is hereby incorporated by reference in its entirety). Many different codons can be 10 mutated, such as Ser249. p53 cfDNA mutations have been identified in breast, lung, ovarian, bladder, gastric, pancreatic, colorectal, bowel, and hepatocellular cancers (Fleischhacker & Schmidt "Circulating nucleic acids (CNAs) and caner—a survey," Biochim Biophys Acta 1775:181-232, 15 2007, which is hereby incorporated by reference in its

BRAF is an oncogene downstream of Ras. BRAF mutations have been identified in glial neoplasm, melanoma, thyroid, and lung cancers (Dias-Santagata et al. BRAF 20 V600E mutations are common in pleomorphic xanthoastrocytoma: diagnostic and therapeutic implications. PLOS ONE 2011; 6:e17948, 2011; Shinozaki et al. Utility of circulating B-RAF DNA mutation in serum for monitoring melanoma patients receiving biochemotherapy. Clin Canc 25 Res 13:2068-2074, 2007; and Board et al. Detection of BRAF mutations in the tumor and serum of patients enrolled in the AZD6244 (ARRY-142886) advanced melanoma phase II study. Brit J Canc 2009; 101:1724-1730, which are each hereby incorporated by reference in its entirety). The 30 BRAF V600E mutation occurs, e.g., in melanoma tumors, and is more common in advanced stages. The V600E mutation has been detected in cfDNA.

EGFR contributes to cell proliferation and is misregulated in many cancers (Downward J. Targeting RAS signaling 35 pathways in cancer therapy. Nature Rev Cancer 3:11-22, 2003; and Levine & Oren "The first 30 years of p53: growing ever more complex. Nature Rev Cancer," 9:749-758, 2009, which is hereby incorporated by reference in its entirety). Exemplary EGFR mutations include those in 40 exons 18-21, which have been identified in lung cancer patients. EGFR cfDNA mutations have been identified in lung cancer patients (Jia et al. "Prediction of epidermal growth factor receptor mutations in the plasma/pleural effusion to efficacy of gefitinib treatment in advanced non-small 45 cell lung cancer," J Canc Res Clin Oncol 2010; 136:1341-1347, 2010, which is hereby incorporated by reference in its

Exemplary polymorphisms or mutations associated with breast cancer include LOH at microsatellites (Kohler et al. 50 "Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors,' Mol Cancer 8:doi:10.1186/1476-4598-8-105, 2009, which is hereby incorporated by reference in its entirety), p53 mutations (such as mutations in exons 5-8)(Garcia et al." Extra- 55 cellular tumor DNA in plasma and overall survival in breast cancer patients," Genes, Chromosomes & Cancer 45:692-701, 2006, which is hereby incorporated by reference in its entirety), HER2 (Sorensen et al. "Circulating HER2 DNA after trastuzumab treatment predicts survival and response in 60 breast cancer," Anticancer Res30:2463-2468, 2010, which is hereby incorporated by reference in its entirety), PIK3CA, MED1, and GAS6 polymorphisms or mutations (Murtaza et al. "Non-invasive analysis of acquired resistance to cancer 10.1038/nature12065, 2013, which is hereby incorporated by reference in its entirety).

134

Increased cfDNA levels and LOH are associated with decreased overall and disease-free survival. p53 mutations (exons 5-8) are associated with decreased overall survival. Decreased circulating HER2 cfDNA levels are associated with a better response to HER2-targeted treatment in HER2positive breast tumor subjects. An activating mutation in PIK3CA, a truncation of MED1, and a splicing mutation in GAS6 result in resistance to treatment.

Exemplary polymorphisms or mutations associated with colorectal cancer include p53, APC, K-ras, and thymidylate synthase mutations and p16 gene methylation (Wang et al. "Molecular detection of APC, K-ras, and p53 mutations in the serum of colorectal cancer patients as circulating biomarkers," World J Surg 28:721-726, 2004; Ryan et al. "A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up," Gut 52:101-108, 2003; Lecomte et al. "Detection of free-circulating tumorassociated DNA in plasma of colorectal cancer patients and its association with prognosis," Int J Cancer 100:542-548, 2002; Schwarzenbach et al. "Molecular analysis of the polymorphisms of thymidylate synthase on cell-free circulating DNA in blood of patients with advanced colorectal carcinoma," Int J Cancer 127:881-888, 2009, which are each hereby incorporated by reference in its entirety). Postoperative detection of K-ras mutations in serum is a strong predictor of disease recurrence. Detection of K-ras mutations and p16 gene methylation are associated with decreased survival and increased disease recurrence. Detection of K-ras, APC, and/or p53 mutations is associated with recurrence and/or metastases. Polymorphisms (including LOH, SNPs, variable number tandem repeats, and deletion) in the thymidylate synthase (the target of fluoropyrimidinebased chemotherapies) gene using cfDNA may be associated with treatment response.

Exemplary polymorphisms or mutations associated with lung cancer (such as non-small cell lung cancer) include K-ras (such as mutations in codon 12) and EGFR mutations. Exemplary prognostic mutations include EGFR mutations (exon 19 deletion or exon 21 mutation) associated with increased overall and progression-free survival and K-ras mutations (in codons 12 and 13) are associated with decreased progression-free survival (Jian et al. "Prediction of epidermal growth factor receptor mutations in the plasma/ pleural effusion to efficacy of gefitinib treatment in advanced non-small cell lung cancer," J Canc Res Clin Oncol 136: 1341-1347, 2010; Wang et al. "Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer," Clin Canc Res 16:1324-1330, 2010, which are each hereby incorporated by reference in its entirety). Exemplary polymorphisms or mutations indicative of response to treatment include EGFR mutations (exon 19 deletion or exon 21 mutation) that improve response to treatment and K-ras mutations (codons 12 and 13) that decrease the response to treatment. A resistance-conferring mutation in EFGR has been identified (Murtaza et al. "Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA, Nature doi:10.1038/nature12065, 2013, which is hereby incorporated by reference in its entirety).

Exemplary polymorphisms or mutations associated with melanoma (such as uveal melanoma) include those in GNAQ, GNA11, BRAF, and p53. Exemplary GNAQ and GNA11 mutations include R183 and Q209 mutations. Q209 therapy by sequencing of plasma DNA," Nature 2013;doi: 65 mutations in GNAQ or GNA11 are associated with metastases to bone. BRAF V600E mutations can be detected in patients with metastatic/advanced stage melanoma. BRAF

## US 11,530,454 B2

135

 $V600\mathrm{E}$  is an indicator of invasive melanoma. The presence of the BRAF V600E mutation after chemotherapy is associated with a non-response to the treatment

Exemplary polymorphisms or mutations associated with pancreatic carcinomas include those in K-ras and p53 (such as p53 Ser249). p53 Ser249 is also associated with hepatitis B infection and hepatocellular carcinoma, as well as ovarian cancer, and non-Hodgkin's lymphoma.

Even polymorphisms or mutations that are present in low frequency in a sample can be detected with the methods of 10 the invention. For example, a polymorphism or mutation that is present at a frequency of 1 in a million can be observed 10 times by performing 10 million sequencing reads. If desired, the number of sequencing reads can be altered depending of the level of sensitivity desired. In some 15 embodiments, a sample is re-analyzed or another sample from a subject is analyzed using a greater number of sequencing reads to improve the sensitivity. For example, if no or only a small number (such as 1, 2, 3, 4, or 5) polymorphisms or mutations that are associated with cancer or an increased risk for cancer are detected, the sample is re-analyzed or another sample is tested.

In some embodiments, multiple polymorphisms or mutations are required for cancer or for metastatic cancer. In such cases, screening for multiple polymorphisms or mutations 25 improves the ability to accurately diagnose cancer or metastatic cancer. In some embodiments when a subject has a subset of multiple polymorphisms or mutations that are required for cancer or for metastatic cancer, the subject can be re-screened later to see if the subject acquires additional 30 mutations

In some embodiments in which multiple polymorphisms or mutations are required for cancer or for metastatic cancer, the frequency of each polymorphism or mutation can be compared to see if they occur at similar frequencies. For 35 example, if two mutations required for cancer (denoted "A" and "B"), some cells will have none, some cells with A, some with B, and some with A and B. If A and B are observed at similar frequencies, the subject is more likely to have some cells with both A and B. If observer A and B at dissimilar frequencies, the subject is more likely to have different cell populations.

In some embodiments in which multiple polymorphisms or mutations are required for cancer or for metastatic cancer, the number or identity of such polymorphisms or mutations 45 that are present in the subject can be used to predict how likely or soon the subject is likely to have the disease or disorder. In some embodiments in which polymorphisms or mutations tend to occur in a certain order, the subject may be periodically tested to see if the subject has acquired the 50 other polymorphisms or mutations.

In some embodiments, determining the presence or absence of multiple polymorphisms or mutations (such as 2, 3, 4, 5, 8, 10, 12, 15, or more) increases the sensitivity and/or specificity of the determination of the presence or absence of 55 a disease or disorder such as cancer, or an increased risk for with a disease or disorder such as cancer.

In some embodiments, the polymorphism(s) or mutation(s) are directly detected. In some embodiments, the polymorphism(s) or mutation(s) are indirectly detected by 60 detection of one or more sequences (e.g., a polymorphic locus such as a SNP) that are linked to the polymorphism or mutation.

**Exemplary Nucleic Acid Alterations** 

In some embodiments, there is a change to the integrity of 65 RNA or DNA (such as a change in the size of fragmented cfRNA or cfDNA or a change in nucleosome composition)

136

that is associated with a disease or disorder such as cancer, or an increased risk for a disease or disorder such as cancer. In some embodiments, there is a change in the methylation pattern RNA or DNA that is associated with a disease or disorder such as cancer, or an increased risk for with a disease or disorder such as cancer (e.g., hypermethylation of tumor suppressor genes). For example, methylation of the CpG islands in the promoter region of tumor-suppressor genes has been suggested to trigger local gene silencing. Aberrant methylation of the p16 tumor suppressor gene occurs in subjects with liver, lung, and breast cancer. Other frequently methylated tumor suppressor genes, including APC, Ras association domain family protein 1A (RASSF1A), glutathione S-transferase P1 (GSTP1), and DAPK, have been detected in various type of cancers, for example nasopharyngeal carcinoma, colorectal cancer, lung cancer, esophageal cancer, prostate cancer, bladder cancer, melanoma, and acute leukemia. Methylation of certain tumor-suppressor genes, such as p16, has been described as an early event in cancer formation, and thus is useful for early cancer screening.

In some embodiments, bisulphite conversion or a non-bisulphite based strategy using methylation sensitive restriction enzyme digestion is used to determine the methylation pattern (Hung et al., J Clin Pathol 62:308-313, 2009, which is hereby incorporated by reference in its entirety). On bisulphite conversion, methylated cytosines remain as cytosines while unmethylated cytosines are converted to uracils. Methylation-sensitive restriction enzymes (e.g., BstUI) cleaves unmethylated DNA sequences at specific recognition sites (e.g., 5'-CG \rangle CG-3' for BstUI), while methylated sequences remain intact. In some embodiments, the intact methylated sequences are detected. In some embodiments, stem-loop primers are used to selectively amplify restriction enzyme-digested unmethylated fragments without co-amplifying the non-enzyme-digested methylated DNA.

Exemplary Changes in mRNA Splicing

In some embodiments, a change in mRNA splicing is associated with a disease or disorder such as cancer, or an increased risk for a disease or disorder such as cancer. In some embodiments, the change in mRNA splicing is in one or more of the following nucleic acids associated with cancer or an increased risk for cancer: DNMT3B, BRCA1, KLF6, Ron, or Gemin5. In some embodiments, the detected mRNA splice variant is associated with a disease or disorder, such as cancer. In some embodiments, multiple mRNA splice variants are produced by healthy cells (such as non-cancerous cells), but a change in the relative amounts of the mRNA splice variants is associated with a disease or disorder, such as cancer. In some embodiments, the change in mRNA splicing is due to a change in the mRNA sequence (such as a mutation in a splice site), a change in splicing factor levels, a change in the amount of available splicing factor (such as a decrease in the amount of available splicing factor due to the binding of a splicing factor to a repeat), altered splicing regulation, or the tumor microenvironment.

The splicing reaction is carried out by a multi-protein/RNA complex called the spliceosome (Fackenthal1 and Godley, Disease Models & Mechanisms 1: 37-42, 2008, doi:10.1242/dmm.000331, which is hereby incorporated by reference in its entirety). The spliceosome recognizes intronexon boundaries and removes intervening introns via two transesterification reactions that result in ligation of two adjacent exons. The fidelity of this reaction must be exquisite, because if the ligation occurs incorrectly, normal protein-encoding potential may be compromised. For example, in cases where exon-skipping preserves the reading frame of

## US 11,530,454 B2

137

the triplet codons specifying the identity and order of amino acids during translation, the alternatively spliced mRNA may specify a protein that lacks crucial amino acid residues. More commonly, exon-skipping will disrupt the translational reading frame, resulting in premature stop codons. 5 These mRNAs are typically degraded by at least 90% through a process known as nonsense-mediated mRNA degradation, which reduces the likelihood that such defective messages will accumulate to generate truncated protein products. If mis-spliced mRNAs escape this pathway, then 10 truncated, mutated, or unstable proteins are produced.

Alternative splicing is a means of expressing several or many different transcripts from the same genomic DNA and results from the inclusion of a subset of the available exons for a particular protein. By excluding one or more exons, 15 certain protein domains may be lost from the encoded protein, which can result in protein function loss or gain. Several types of alternative splicing have been described: exon skipping; alternative 5' or 3' splice sites; mutually exclusive exons; and, much more rarely, intron retention. 20 Others have compared the amount of alternative splicing in cancer versus normal cells using a bioinformatic approach and determined that cancers exhibit lower levels of alternative splicing than normal cells. Furthermore, the distribution of the types of alternative splicing events differed in cancer 25 versus normal cells. Cancer cells demonstrated less exon skipping, but more alternative 5' and 3' splice site selection and intron retention than normal cells. When the phenomenon of exonization (the use of sequences as exons that are used predominantly by other tissues as introns) was exam- 30 ined, genes associated with exonization in cancer cells were preferentially associated with mRNA processing, indicating a direct link between cancer cells and the generation of aberrant mRNA splice forms.

Exemplary Changes in DNA or RNA Levels

In some embodiments, there is a change in the total amount or concentration of one or more types of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic 40 RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA). In some embodiments, there is a change in the amount or concentration of one or more specific DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic 45 RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) molecules. In some embodiments, one allele is expressed more than another allele of a locus of interest. Exemplary miRNAs are short 20-22 nucleotide RNA mol- 50 ecules that regulate the expression of a gene. In some embodiments, there is a change in the transcriptome, such as a change in the identity or amount of one or more RNA molecules.

In some embodiments, an increase in the total amount or concentration of cfDNA or cfRNA is associated with a disease or disorder such as cancer, or an increased risk for a disease or disorder such as cancer. In some embodiments, the total concentration of a type of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) increases by at least 2, 3, 4, 5, 6, 7, 8, 9, 10-fold, or more compared to the total concentration of that type of DNA or RNA in 65 healthy (such as non-cancerous) subjects. In some embodiments, a total concentration of cfDNA between 75 to 100

138

ng/mL, 100 to 150 ng/mL, 150 to 200 ng/mL, 200 to 300 ng/mL, 300 to 400 ng/mgL, 400 to 600 ng/mL, 600 to 800 ng/mL, 800 to 1,000 ng/mL, inclusive, or a total concentration of cfDNA of more than 100 ng, mL, such as more than 200, 300, 400, 500, 600, 700, 800, 900, or 1,000 ng/mL is indicative of cancer, an increased risk for cancer, an increased risk of a tumor being malignant rather than benign, a decreased probably of the cancer going into remission, or a worse prognosis for the cancer. In some embodiments, the amount of a type of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) having one or more polymorphisms/mutations (such as deletions or duplications) associated with a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25% of the total amount of that type of DNA or RNA. In some embodiments, at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25% of the total amount of a type of DNA (such as cfDNA cf mDNA, cf nDNA, cellular DNA, or mitochondrial DNA) or RNA (cfRNA, cellular RNA, cytoplasmic RNA, coding cytoplasmic RNA, non-coding cytoplasmic RNA, mRNA, miRNA, mitochondrial RNA, rRNA, or tRNA) has a particular polymorphism or mutation (such as a deletion or duplication) associated with a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer.

In some embodiments, the cfDNA is encapsulated. In some embodiments, the cfDNA is not encapsulated.

In some embodiments, the fraction of tumor DNA out of total DNA (such as fraction of tumor cfDNA out of total cfDNA or fraction of tumor cfDNA with a particular mutation out of total cfDNA) is determined. In some embodiments, the fraction of tumor DNA may be determined for a plurality of mutations, where the mutations can be single nucleotide variants, copy number variants, differential methylation, or combinations thereof. In some embodiments, the average tumor fraction calculated for one or a set of mutations with the highest calculated tumor fraction is taken as the actual tumor fraction in the sample. In some embodiments, the average tumor fraction calculated for all of the mutations is taken as the actual tumor fraction in the sample. In some embodiments, this tumor fraction is used to stage a cancer (since higher tumor fractions can be associated with more advanced stages of cancer). In some embodiments, the tumor fraction is used to size a cancer, since larger tumors may be correlated with the fraction of tumor DNA in the plasma. In some embodiments, the tumor fraction is used to size the proportion of a tumor that is afflicted with a single or plurality of mutations, since there may be a correlation between the measured tumor fraction in a plasma sample and the size of tissue with a given mutation(s) genotype. For example, the size of tissue with a given mutation(s) genotype may be correlated with the fraction of tumor DNA that may be calculated by focusing on that particular mutation(s). Exemplary Databases

The invention also features databases containing one or more results from a method of the invention. For example, the database may include records with any of the following information for one or more subjects: any polymorphisms/mutations (such as CNVs) identified, any known association of the polymorphisms/mutations with a disease or disorder or an increased risk for a disease or disorder, effect of the polymorphisms/mutations on the expression or activity level of the encoded mRNA or protein, fraction of DNA, RNA, or

## US 11,530,454 B2

139

cells associated with a disease or disorder (such as DNA, RNA, or cells having polymorphism/mutation associated with a disease or disorder) out of the total DNA, RNA, or cells in sample, source of sample used to identify the polymorphisms/mutations (such as a blood sample or sample from a particular tissue), number of diseased cells, results from later repeating the test (such as repeating the test to monitor the progression or remission of the disease or disorder), results of other tests for the disease or disorder, type of disease or disorder the subject was diagnosed with, 10 treatment(s) administered, response to such treatment(s), side-effects of such treatment(s), symptoms (such as symptoms associated with the disease or disorder), length and number of remissions, length of survival (such as length of time from initial test until death or length of time from 15 diagnosis until death), cause of death, and combinations

In some embodiments, the database includes records with any of the following information for one or more subjects: any polymorphisms/mutations identified, any known asso-20 ciation of the polymorphisms/mutations with cancer or an increased risk for cancer, effect of the polymorphisms/ mutations on the expression or activity level of the encoded mRNA or protein, fraction of cancerous DNA, RNA or cells out of the total DNA, RNA, or cells in sample, source of 25 sample used to identify the polymorphisms/mutations (such as a blood sample or sample from a particular tissue), number of cancerous cells, size of tumor(s), results from later repeating the test (such as repeating the test to monitor the progression or remission of the cancer), results of other 30 tests for cancer, type of cancer the subject was diagnosed with, treatment(s) administered, response to such treatment(s), side-effects of such treatment(s), symptoms (such as symptoms associated with cancer), length and number of remissions, length of survival (such as length of time from 3: initial test until death or length of time from cancer diagnosis until death), cause of death, and combinations thereof. In some embodiments, the response to treatment includes any of the following: reducing or stabilizing the size of a tumor (e.g., a benign or cancerous tumor), slowing or preventing 40 an increase in the size of a tumor, reducing or stabilizing the number of tumor cells, increasing the disease-free survival time between the disappearance of a tumor and its reappearance, preventing an initial or subsequent occurrence of a tumor, reducing or stabilizing an adverse symptom associ- 45 ated with a tumor, or combinations thereof. In some embodiments, the results from one or more other tests for a disease or disorder such as cancer are included, such as results from screening tests, medical imaging, or microscopic examination of a tissue sample.

In one such aspect, the invention features an electronic database including at least 5, 10,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$  or more records. In some embodiments, the database has records for at least 5, 10,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$  or more different subjects.

In another aspect, the invention features a computer including a database of the invention and a user interface. In some embodiments, the user interface is capable of displaying a portion or all of the information contained in one or more records. In some embodiments, the user interface is capable of displaying (i) one or more types of cancer that have been identified as containing a polymorphism or mutation whose record is stored in the computer, (ii) one or more polymorphisms or mutations that have been identified in a particular type of cancer whose record is stored in the computer, (iii) prognosis information for a particular type of cancer or a particular a polymorphism or mutation whose

140

record is stored in the computer (iv) one or more compounds or other treatments useful for cancer with a polymorphism or mutation whose record is stored in the computer, (v) one or more compounds that modulate the expression or activity of an mRNA or protein whose record is stored in the computer, and (vi) one or more mRNA molecules or proteins whose expression or activity is modulated by a compound whose record is stored in the computer. The internal components of the computer typically include a processor coupled to a memory. The external components usually include a massstorage device, e.g., a hard disk drive; user input devices, e.g., a keyboard and a mouse; a display, e.g., a monitor; and optionally, a network link capable of connecting the computer system to other computers to allow sharing of data and processing tasks. Programs may be loaded into the memory of this system during operation.

In another aspect, the invention features a computerimplemented process that includes one or more steps of any of the methods of the invention.

Exemplary Risk Factors

In some embodiments, the subject is also evaluated for one or more risk factors for a disease or disorder, such as cancer. Exemplary risk factors include family history for the disease or disorder, lifestyle (such as smoking and exposure to carcinogens) and the level of one or more hormones or serum proteins (such as alpha-fetoprotein (AFP) in liver cancer, carcinoembryonic antigen (CEA) in colorectal cancer, or prostate-specific antigen (PSA) in prostate cancer). In some embodiments, the size and/or number of tumors is measured and use in determining a subject's prognosis or selecting a treatment for the subject.

Exemplary Screening Methods

If desired, the presence or absence of a disease or disorder such cancer can be confirmed, or the disease or disorder such as cancer can be classified using any standard method. For example, a disease or disorder such as cancer can be detected in a number of ways, including the presence of certain signs and symptoms, tumor biopsy, screening tests, or medical imaging (such as a mammogram or an ultrasound). Once a possible cancer is detected, it may be diagnosed by microscopic examination of a tissue sample. In some embodiments, a subject diagnosed undergoes repeat testing using a method of the invention or known testing for the disease or disorder at multiple time points to monitor the progression of the disease or disorder or the remission or reoccurrence of the disease or disorder.

**Exemplary Cancers** Exemplary cancers that can be diagnosed, prognosed, stabilized, treated, or prevented using any of the methods of the invention include solid tumors, carcinomas, sarcomas, lymphomas, leukemias, germ cell tumors, or blastomas. In various embodiments, the cancer is an acute lymphoblastic leukemia, acute myeloid leukemia, adrenocortical carcinoma, AIDS-related cancer, AIDS-related lymphoma, anal cancer, appendix cancer, astrocytoma (such as childhood cerebellar or cerebral astrocytoma), basal-cell carcinoma, bile duct cancer (such as extrahepatic bile duct cancer) bladder cancer, bone tumor (such as osteosarcoma or malignant fibrous histiocytoma), brainstem glioma, brain cancer (such as cerebellar astrocytoma, cerebral astrocytoma/malignant glioma, ependymo, medulloblastoma, supratentorial primitive neuroectodermal tumors, or visual pathway and hypothalamic glioma), glioblastoma, breast cancer, bronchial adenoma or carcinoid, burkitt's lymphoma, carcinoid tumor (such as a childhood or gastrointestinal carcinoid tumor), carcinoma central nervous system lymphoma, cerebellar astrocytoma or malignant glioma (such as childhood

## US 11,530,454 B2

141

cerebellar astrocytoma or malignant glioma), cervical cancer, childhood cancer, chronic lymphocytic leukemia, chronic myelogenous leukemia, chronic myeloproliferative disorders, colon cancer, cutaneous t-cell lymphoma, desmoplastic small round cell tumor, endometrial cancer, ependymoma, esophageal cancer, ewing's sarcoma, tumor in the ewing family of tumors, extracranial germ cell tumor (such as a childhood extracranial germ cell tumor), extragonadal germ cell tumor, eye cancer (such as intraocular melanoma or retinoblastoma eye cancer), gallbladder can- 10 cer, gastric cancer, gastrointestinal carcinoid tumor, gastrointestinal stromal tumor, germ cell tumor (such as extracranial, extragonadal, or ovarian germ cell tumor), gestational trophoblastic tumor, glioma (such as brain stem, childhood cerebral astrocytoma, or childhood visual pathway and 15 hypothalamic glioma), gastric carcinoid, hairy cell leukemia, head and neck cancer, heart cancer, hepatocellular (liver) cancer, hodgkin lymphoma, hypopharyngeal cancer, hypothalamic and visual pathway glioma (such as childhood visual pathway glioma), islet cell carcinoma (such as endo-20 crine or pancreas islet cell carcinoma), kaposi sarcoma, kidney cancer, laryngeal cancer, leukemia (such as acute lymphoblastic, acute myeloid, chronic lymphocytic, chronic myelogenous, or hairy cell leukemia), lip or oral cavity cancer, liposarcoma, liver cancer (such as non-small cell or 25 small cell cancer), lung cancer, lymphoma (such as AIDSrelated, burkitt, cutaneous T cell, Hodgkin, non-hodgkin, or central nervous system lymphoma), macroglobulinemia (such as waldenström macroglobulinemia, malignant fibrous histiocytoma of bone or osteosarcoma, medulloblastoma 30 (such as childhood medulloblastoma), melanoma, merkel cell carcinoma, mesothelioma (such as adult or childhood mesothelioma), metastatic squamous neck cancer with occult, mouth cancer, multiple endocrine neoplasia syndrome (such as childhood multiple endocrine neoplasia 35 syndrome), multiple myeloma or plasma cell neoplasm. mycosis fungoides, myelodysplastic syndrome, myelodysplastic or myeloproliferative disease, myelogenous leukemia (such as chronic myelogenous leukemia), myeloid leukemia (such as adult acute or childhood acute myeloid 40 leukemia), myeloproliferative disorder (such as chronic myeloproliferative disorder), nasal cavity or paranasal sinus cancer, nasopharyngeal carcinoma, neuroblastoma, oral cancer, oropharyngeal cancer, osteosarcoma or malignant fibrous histiocytoma of bone, ovarian cancer, ovarian epi- 45 thelial cancer, ovarian germ cell tumor, ovarian low malignant potential tumor, pancreatic cancer (such as islet cell pancreatic cancer), paranasal sinus or nasal cavity cancer, parathyroid cancer, penile cancer, pharyngeal cancer, pheochromocytoma, pineal astrocytoma, pineal germinoma. 50 pineoblastoma or supratentorial primitive neuroectodermal tumor (such as childhood pineoblastoma or supratentorial primitive neuroectodermal tumor), pituitary adenoma, plasma cell neoplasia, pleuropulmonary blastoma, primary central nervous system lymphoma, cancer, rectal cancer, 55 renal cell carcinoma, renal pelvis or ureter cancer (such as renal pelvis or ureter transitional cell cancer, retinoblastoma, rhabdomyosarcoma (such as childhood rhabdomyosarcoma), salivary gland cancer, sarcoma (such as sarcoma in the ewing family of tumors, Kaposi, soft tissue, or uterine 60 sarcoma), sézary syndrome, skin cancer (such as nonmelanoma, melanoma, or merkel cell skin cancer), small intestine cancer, squamous cell carcinoma, supratentorial primitive neuroectodermal tumor (such as childhood supratentorial primitive neuroectodermal tumor), T-cell lymphoma (such 65 as cutaneous T-cell lymphoma), testicular cancer, throat cancer, thymoma (such as childhood thymoma), thymoma or

142

thymic carcinoma, thyroid cancer (such as childhood thyroid cancer), trophoblastic tumor (such as gestational trophoblastic tumor), unknown primary site carcinoma (such as adult or childhood unknown primary site carcinoma), urethral cancer (such as endometrial uterine cancer), uterine sarcoma, vaginal cancer, visual pathway or hypothalamic glioma (such as childhood visual pathway or hypothalamic glioma), vulvar cancer, waldenström macroglobulinemia, or wilms tumor (such as childhood wilms tumor). In various embodiments, the cancer has metastasized or has not metastasized.

The cancer may or may not be a hormone related or dependent cancer (e.g., an estrogen or androgen related cancer). Benign tumors or malignant tumors may be diagnosed, prognosed, stabilized, treated, or prevented using the methods and/or compositions of the present invention.

In some embodiments, the subject has a cancer syndrome. A cancer syndrome is a genetic disorder in which genetic mutations in one or more genes predispose the affected individuals to the development of cancers and may also cause the early onset of these cancers. Cancer syndromes often show not only a high lifetime risk of developing cancer, but also the development of multiple independent primary tumors. Many of these syndromes are caused by mutations in tumor suppressor genes, genes that are involved in protecting the cell from turning cancerous. Other genes that may be affected are DNA repair genes, oncogenes and genes involved in the production of blood vessels (angiogenesis). Common examples of inherited cancer syndromes are hereditary breast-ovarian cancer syndrome and hereditary non-polyposis colon cancer (Lynch syndrome).

In some embodiments, a subject with one or more polymorphisms or mutations n K-ras, p53, BRA, EGFR, or HER2 is administered a treatment that targets K-ras, p53, BRA, EGFR, or HER2, respectively.

The methods of the invention can be generally applied to the treatment of malignant or benign tumors of any cell, tissue, or organ type.

**Exemplary Treatments** 

If desired, any treatment for stabilizing, treating, or preventing a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer can be administered to a subject (e.g., a subject identified as having cancer or an increased risk for cancer using any of the methods of the invention). In various embodiments, the treatment is a known treatment or combination of treatments for a disease or disorder such as cancer, such as cytotoxic agents, targeted therapy, immunotherapy, hormonal therapy, radiation therapy, surgical removal of cancerous cells or cells likely to become cancerous, stem cell transplantation, bone marrow transplantation, photodynamic therapy, palliative treatment, or a combination thereof. In some embodiments, a treatment (such as a preventative medication) is used to prevent, delay, or reduce the severity of a disease or disorder such as cancer in a subject at increased risk for a disease or disorder such as cancer.

In some embodiments, the targeted therapy is a treatment that targets the cancer's specific genes, proteins, or the tissue environment that contributes to cancer growth and survival. This type of treatment blocks the growth and spread of cancer cells while limiting damage to normal cells, usually leading to fewer side effects than other cancer medications.

One of the more successful approaches has been to target angiogenesis, the new blood vessel growth around a tumor. Targeted therapies such as bevacizumab (Avastin), lenalidomide (Revlimid), sorafenib (Nexavar), sunitinib (Sutent), and thalidomide (Thalomid) interfere with angiogenesis.

## US 11,530,454 B2

143

Another example is the use of a treatment that targets HER2, such as trastuzumab or lapatinib, for cancers that overexpress HER2 (such as some breast cancers). In some embodiments, a monoclonal antibody is used to block a specific target on the outside of cancer cells. Examples include alemtuzumab (Campath-1H), bevacizumab, cetuximab (Erbitux), panitumumab (Vectibix), pertuzumab (Omnitarg), rituximab (Rituxan), and trastuzumab. In some embodiments, the monoclonal antibody tositumomab (Bexxar) is used to deliver radiation to the tumor. In some embodiments, 10 an oral small molecule inhibits a cancer process inside of a cancer cell. Examples include dasatinib (Sprycel), erlotinib (Tarceva), gefitinib (Iressa), imatinib (Gleevec), lapatinib (Tykerb), nilotinib (Tasigna), sorafenib, sunitinib, and temsirolimus (Torisel). In some embodiments, a proteasome 15 inhibitor (such as the multiple myeloma drug, bortezomib (Velcade)) interferes with specialized proteins called enzymes that break down other proteins in the cell-

In some embodiments, immunotherapy is designed to boost the body's natural defenses to fight the cancer. Exemplary types of immunotherapy use materials made either by the body or in a laboratory to bolster, target, or restore immune system function.

In some embodiments, hormonal therapy treats cancer by lowering the amounts of hormones in the body. Several 25 types of cancer, including some breast and prostate cancers, only grow and spread in the presence of natural chemicals in the body called hormones. In various embodiments, hormonal therapy is used to treat cancers of the prostate, breast, thyroid, and reproductive system.

In some embodiments, the treatment includes a stem cell transplant in which diseased bone marrow is replaced by highly specialized cells, called hematopoietic stem cells. Hematopoietic stem cells are found both in the bloodstream and in the bone marrow.

In some embodiments, the treatment includes photodynamic therapy, which uses special drugs, called photosensitizing agents, along with light to kill cancer cells. The drugs work after they have been activated by certain kinds of light.

In some embodiments, the treatment includes surgical removal of cancerous cells or cells likely to become cancerous (such as a lumpectomy or a mastectomy). For example, a woman with a breast cancer susceptibility gene mutation (BRCA1 or BRCA2 gene mutation) may reduce 45 her risk of breast and ovarian cancer with a risk reducing salpingo-oophorectomy (removal of the fallopian tubes and ovaries) and/or a risk reducing bilateral mastectomy (removal of both breasts). Lasers, which are very powerful, precise beams of light, can be used instead of blades 50 (scalpels) for very careful surgical work, including treating some cancers.

In addition to treatment to slow, stop, or eliminate the cancer (also called disease-directed treatment), an important part of cancer care is relieving a subject's symptoms and side effects, such as pain and nausea. It includes supporting the subject with physical, emotional, and social needs, an approach called palliative or supportive care. People often receive disease-directed therapy and treatment to ease symptoms at the same time.

Exemplary treatments include actinomycin D, adcetris, Adriamycin, aldesleukin, alemtuzumab, alimta, amsidine, amsacrine, anastrozole, aredia, arimidex, aromasin, asparaginase, avastin, bevacizumab, bicalutamide, bleomycin, bondronat, bonefos, bortezomib, busilvex, busulphan, 65 campto, capecitabine, carboplatin, carmustine, casodex, cetuximab, chimax, chlorambucil, cimetidine, cisplatin,

144

cladribine, clodronate, clofarabine, crisantaspase, cyclophosphamide, cyproterone acetate, cyprostat, cytarabine, cytoxan, dacarbozine, dactinomycin, dasatinib, daunorubicin, dexamethasone, diethylstilbestrol, docetaxel, doxorubicin, drogenil, emcyt, epirubicin, eposin, Erbitux, erlotinib, estracyte, estramustine, etopophos, etoposide, evoltra, exemestane, fareston, femara, filgrastim, fludara, fludarabine, fluorouracil, flutamide, gefinitib, gemcitabine, gemzar, gleevec, glivec. gonapeptyl depot, goserelin, halaven, herceptin, hycamptin, hydroxycarbamide, ibandronic acid, ibritumomab, idarubicin, ifosfomide, interferon, imatinib mesylate, iressa, irinotecan, jevtana, lanvis, lapatinib, letrozole, leukeran, leuprorelin, leustat, lomustine, mabcampath, mabthera, megace, megestrol, methotrexate, mitozantrone, mitomycin, mutulane, myleran, navelbine, neulasta, neupogen, nexavar, nipent, nolvadex D, novantron, oncovin, paclitaxel, pamidronate, PCV, pemetrexed, pentostatin, perj eta, procarbazine, provenge, prednisolone, prostrap, raltitrexed, rituximab, sprycel, sorafenib, soltamox, streptozocin, stilboestrol, stimuvax, sunitinib, sutent, tabloid, tagamet, tamofen, tamoxifen, tarceva, taxol, taxotere, tegafur with uracil, temodal, temozolomide, thalidomide, thioplex, thiotioguanine, tomudex, topotecan, toremifene, trastuzumab, tretinoin, treosulfan, triethylenethiophorsphoramide, triptorelin, tyverb, uftoral, velcade, vepesid, vesanoid, vincristine, vinorelbine, xalkori, xeloda, yervoy, zactima, zanosar, zavedos, zevelin, zoladex, zoledronate, zometa zoledronic acid, and zytiga.

For subjects that express both a mutant form (e.g., a cancer-related form) and a wild-type form (e.g., a form not associated with cancer) of an mRNA or protein, the therapy preferably inhibits the expression or activity of the mutant form by at least 2, 5, 10, or 20-fold more than it inhibits the expression or activity of the wild-type form. The simultaneous or sequential use of multiple therapeutic agents may greatly reduce the incidence of cancer and reduce the number of treated cancers that become resistant to therapy. In addition, therapeutic agents that are used as part of a combination therapy may require a lower dose to treat cancer than the corresponding dose required when the therapeutic agents are used individually. The low dose of each compound in the combination therapy reduces the severity of potential adverse side-effects from the compounds.

In some embodiments, a subject identified as having an increased risk of cancer may invention or any standard method), avoid specific risk factors, or make lifestyle changes to reduce any additional risk of cancer.

In some embodiments, the polymorphisms, mutations, risk factors, or any combination thereof are used to select a treatment regimen for the subject. In some embodiments, a larger dose or greater number of treatments is selected for a subject at greater risk of cancer or with a worse prognosis. Other Compounds for Inclusion in Individual or Combination Therapies

If desired, additional compounds for stabilizing, treating, or preventing a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer may be identified from large libraries of both natural product or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field or drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the methods of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened for their effect on cells from a particular type of cancer or from a particular subject or screened for their

## US 11,530,454 B2

145

effect on the activity or expression of cancer related molecules (such as cancer related molecules known to have altered activity or expression in a particular type of cancer). When a crude extract is found to modulate the activity or expression of a cancer related molecule, further fractionation of the positive lead extract may be performed to isolate chemical constituent responsible for the observed effect using methods known in the art.

Exemplary Assays and Animal Models for the Testing of Therapies

If desired, one or more of the treatment disclosed herein can be tested for their effect on a disease or disorder such as cancer using a cell line (such as a cell line with one or more of the mutations identified in the subject who has been diagnosed with cancer or an increased risk of cancer using 15 the methods of the invention) or an animal model of the disease or disorder, such as a SCID mouse model (Jain et al., Tumor Models In Cancer Research, ed. Teicher, Humana Press Inc., Totowa, N.J., pp. 647-671, 2001, which is hereby incorporated by reference in its entirety). Additionally, there 20 are numerous standard assays and animal models that can be used to determine the efficacy of particular therapies for stabilizing, treating, or preventing a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer. Therapies can also be tested in standard human 25 clinical trials

For the selection of a preferred therapy for a particular subject, compounds can be tested for their effect on the expression or activity on one or more genes that are mutated in the subject. For example, the ability of a compound to 30 modulate the expression of particular mRNA molecules or proteins can be detected using standard Northern, Western, or microarray analysis. In some embodiments, one or more compounds are selected that (i) inhibit the expression or activity of mRNA molecules or proteins that promote cancer 3 that are expressed at a higher than normal level or have a higher than normal level of activity in the subject (such as in a sample from the subject) or (ii) promote the expression or activity of mRNA molecules or proteins that inhibit cancer that are expressed at a lower than normal level or 40 have a lower than normal level of activity in the subject. An individual or combination therapy that (i) modulates the greatest number of mRNA molecules or proteins that have mutations associated with cancer in the subject and (ii) modulates the least number of mRNA molecules or proteins 45 that do not have mutations associated with cancer in the subject. In some embodiments, the selected individual or combination therapy has high drug efficacy and produces few, if any, adverse side-effects.

As an alternative to the subject-specific analysis described 50 above, DNA chips can be used to compare the expression of mRNA molecules in a particular type of early or late-stage cancer (e.g., breast cancer cells) to the expression in normal tissue (Marrack et al., Current Opinion in Immunology 12, 206-209, 2000; Harkin, Oncologist. 5:501-507, 2000; Pelizzari et al., Nucleic Acids Res. 28(22):4577-4581, 2000, which are each hereby incorporated by reference in its entirety). Based on this analysis, an individual or combination therapy for subjects with this type of cancer can be selected to modulate the expression of the mRNA or proteins 60 that have altered expression in this type of cancer.

In addition to being used to select a therapy for a particular subject or group of subjects, expression profiling can be used to monitor the changes in mRNA and/or protein expression that occur during treatment. For example, expression profiling can be used to determine whether the expression of cancer related genes has returned to normal levels. If

146

not, the dose of one or more compounds in the therapy can be altered to either increase or decrease the effect of the therapy on the expression levels of the corresponding cancer related gene(s). In addition, this analysis can be used to determine whether a therapy affects the expression of other genes (e.g., genes that are associated with adverse side-effects). If desired, the dose or composition of the therapy can be altered to prevent or reduce undesired side-effects. Exemplary Formulations and Methods of Administration

For stabilizing, treating, or preventing a disease or disorder such as cancer or an increased risk for a disease or disorder such as cancer, a composition may be formulated and administered using any method known to those of skill in the art (see, e.g., U.S. Pat. Nos. 8,389,578 and 8,389,557, which are each hereby incorporated by reference in its entirety). General techniques for formulation and administration are found in "Remington: The Science and Practice of Pharmacy," 21st Edition, Ed. David Troy, 2006, Lippincott Williams & Wilkins, Philadelphia, Pa., which is hereby incorporated by reference in its entirety). Liquids, slurries, tablets, capsules, pills, powders, granules, gels, ointments, suppositories, injections, inhalants, and aerosols are examples of such formulations. By way of example, modified or extended release oral formulation can be prepared using additional methods known in the art. For example, a suitable extended release form of an active ingredient may be a matrix tablet or capsule composition. Suitable matrix forming materials include, for example, waxes (e.g., carnauba, bees wax, paraffin wax, ceresin, shellac wax, fatty acids, and fatty alcohols), oils, hardened oils or fats (e.g., hardened rapeseed oil, castor oil, beef tallow, palm oil, and soya bean oil), and polymers (e.g., hydroxypropyl cellulose, polyvinylpyrrolidone, hydroxypropyl methyl cellulose, and polyethylene glycol). Other suitable matrix tableting materials are microcrystalline cellulose, powdered cellulose, hydroxypropyl cellulose, ethyl cellulose, with other carriers, and fillers. Tablets may also contain granulates, coated powders, or pellets. Tablets may also be multi-layered. Optionally, the finished tablet may be coated or uncoated.

Typical routes of administering such compositions include, without limitation, oral, sublingual, buccal, topical, transdermal, inhalation, parenteral (e.g., subcutaneous, intravenous, intramuscular, intrasternal injection, or infusion techniques), rectal, vaginal, and intranasal. In preferred embodiments, the therapy is administered using an extended release device. Compositions of the invention are formulated so as to allow the active ingredient(s) contained therein to be bioavailable upon administration of the composition. Compositions may take the form of one or more dosage units. Compositions may contain 1, 2, 3, 4, or more active ingredients and may optionally contain 1, 2, 3, 4, or more inactive ingredients.

#### ALTERNATE EMBODIMENTS

Any of the methods described herein may include the output of data in a physical format, such as on a computer screen, or on a paper printout. Any of the methods of the invention may be combined with the output of the actionable data in a format that can be acted upon by a physician. Some of the embodiments described in the document for determining genetic data pertaining to a target individual may be combined with the notification of a potential chromosomal abnormality (such as a deletion or duplication), or lack thereof, with a medical professional, optionally combined with the decision to abort, or to not abort, a fetus in the context of prenatal diagnosis. Some of the embodiments

## US 11,530,454 B2

147

described herein may be combined with the output of the actionable data, and the execution of a clinical decision that results in a clinical treatment, or the execution of a clinical decision to make no action.

In some embodiments, a method is disclosed herein for 5 generating a report disclosing a result of any method of the invention (such as the presence or absence of a deletion or duplication). A report may be generated with a result from a method of the invention, and it may be sent to a physician electronically, displayed on an output device (such as a 10 digital report), or a written report (such as a printed hard copy of the report) may be delivered to the physician. In addition, the described methods may be combined with the actual execution of a clinical decision that results in a clinical treatment, or the execution of a clinical decision to 15 make no action.

In certain embodiments, the present invention provides reagents, kits, and methods, and computer systems and computer media with encoded instructions for performing such methods, for detecting both CNVs and SNVs from the 20 same sample using the multiplex PCR methods disclosed herein. In certain preferred embodiments the sample is a single cell sample or a plasma sample suspected of containing circulating tumor DNA. These embodiments take advantage of the discovery that by interrogating DNA samples 25 from single cells or plasma for CNVs and SNVs using the highly sensitive multiplex PCR methods disclosed herein, improved cancer detection can be achieved, versus interrogating for either CNVs or SNVs alone, especially for cancers exhibiting CNV such as breast, ovarian, and lung 30 cancer. The methods in certain illustrative embodiments for analyzing CNVs interrogate for between 50 and 100,000 or 50 and 10,000, or 50 and 1,000 SNPs and for SNVs interrogate for between 50 and 1000 SNVs or for between 50 and 500 SNVs or for between 50 and 250 SNVs. The 3: methods provided herein for detecting CNVs and/or SNVs in plasma of subjects suspected of having cancer, including for example, cancers known to exhibit CNVs and SNVs, such as breast, lung, and ovarian cancer, provide the advantage of detecting CNVs and/or SNVs from tumors that often 40 are composed of heterogeneous cancer cell populations in terms of genetic compositions. Thus, traditional methods, which focus on analyzing only certain regions of the tumors can often miss CNVs or SNVs that are present in cells in other regions of the tumor. The plasma samples act as liquid 45 biopsies that can be interrogated to detect any of the CNVs and/or SNVs that are present in only subpopulations of

Example Computer Architecture

FIG. **69** shows an example system architecture **X00** useful 50 for performing embodiments of the present invention. System architecture X00 includes an analysis platform X08 connected to one or more laboratory information systems ("LISs") X04. As shown in FIG. 69, analysis platform X08 may be connected to LIS X04 over a network X02. Network 55 X02 may include one or more networks of one or more network types, including any combination of LAN, WAN, the Internet, etc. Network X02 may encompass connections between any or all components in system architecture X00. Analysis platform X08 may alternatively or additionally be 60 connected directly to LIS X06. In an embodiment, analysis platform X08 analyzes genetic data provided by LIS X04 in a software-as-a-service model, where LIS X04 is a thirdparty LIS, while analysis platform X08 analyzes genetic data provided by LIS X06 in a full-service or in-house 65 model, where LIS X06 and analysis platform X08 are controlled by the same party. In an embodiment where

148

analysis platform X08 is providing information over network X02, analysis platform X08 may be a server.

In an example embodiment, laboratory information system X04 includes one or more public or private institutions that collect, manage, and/or store genetic data. A person having skill in the relevant art(s) would understand that methods and standards for securing genetic data are known and can be implemented using various information security techniques and policies, e.g., username/password, Transport Layer Security (TLS), Secure Sockets Layer (SSL), and/or other cryptographic protocols providing communication security.

In an example embodiment, system architecture X00 operates as a service-oriented architecture and uses a clientserver model that would be understood by one of skill in the relevant art(s) to enable various forms of interaction and communication between LIS X04 and analysis platform X08. System architecture X00 may be distributed over various types of networks X02 and/or may operate as cloud computing architecture. Cloud computing architecture may include any type of distributed network architecture. By way of example and not of limitation, cloud computing architecture is useful for providing software as a service (SaaS), infrastructure as a service (IaaS), platform as a service (PaaS), network as a service (NaaS), data as a service (DaaS), database as a service (DBaaS), backend as a service (BaaS), test environment as a service (TEaaS), API as a service (APIaaS), integration platform as a service (IPaaS)

In an example embodiment, LISs X04 and X06 each include a computer, device, interface, etc. or any sub-system thereof. LISs X04 and X06 may include an operating system (OS), applications installed to perform various functions such as, for example, access to and/or navigation of data made accessible locally, in memory, and/or over network X02. In an embodiment, LIS X04 accesses analysis platform X08 through an application programming interface ("API"). LIS X04 may also include one or more native applications that may operate independently of an API.

In an example embodiment, analysis platform X08 includes one or more of an input processor X12, a hypothesis manager X14, a modeler X16, an error correction unit X18, a machine learning unit X20, and an output processor X18. Input processor X12 receives and processes inputs from LISs X04 and/or X06. Processing may include but is not limited to operations such as parsing, transcoding, translating, adapting, or otherwise handling any input received from LISs X04 and/or X06. Inputs may be received via one or more streams, feeds, databases, or other sources of data, such as may be made accessible by LISs X04 and X06. Data errors may be corrected by error correction unit X18 through performance of the error correction mechanisms described above.

In an example embodiment, hypothesis manager X14 is configured to receive the inputs passed from input processor X12 in a form ready to be processed in accordance with hypotheses for genetic analysis that are represented as models and/or algorithms. Such models and/or algorithms may be used by modeler X16 to generate probabilities, for example, based on dynamic, real-time, and/or historical statistics or other indicators. Data used to derive and populate such strategy models and/or algorithms are available to hypothesis manager X14 via, for example, genetic data source X10. Genetic data source X10 may include, for example, a nucleic acid sequencer. Hypothesis manager X14 may be configured to formulate hypotheses based on, for example, the variables required to populate its models

## US 11,530,454 B2

149

and/or algorithms. Models and/or algorithms, once populated, may be used by modeler X16 to generate one or more hypotheses as described above. Hypothesis manager X14 may select a particular value, range of values, or estimate based on a most-likely hypothesis as an output as described above. Modeler X16 may operate in accordance with models and/or algorithms trained by machine learning unit X20. For example, machine learning unit X20 may develop such models and/or algorithms by applying a classification algorithm as described above to a training set database (not shown). In certain embodiments, the machine learning unit analyzes one or more control samples to generate training data sets useful in SNV detections methods provided herein.

Once hypothesis manager X14 has identified a particular output, such output may be returned to the particular LIS 104 or 106 requesting the information by output processor X22.

Various aspects of the disclosure can be implemented on a computing device by software, firmware, hardware, or a combination thereof. FIG. 70 illustrates an example computer system Y00 in which the contemplated embodiments, or portions thereof, can be implemented as computer-readable code. Various embodiments are described in terms of this example computer system Y00.

Processing tasks in the embodiment of FIG. 70 are carried 25 out by one or more processors Y02. However, it should be noted that various types of processing technology may be used here, including programmable logic arrays (PLAs), application-specific integrated circuits (ASICs), multi-core processors, multiple processors, or distributed processors. 30 Additional specialized processing resources such as graphics, multimedia, or mathematical processing capabilities may also be used to aid in certain processing tasks. These processing resources may be hardware, software, or an appropriate combination thereof. For example, one or more 35 of processors Y02 may be a graphics-processing unit (GPU). In an embodiment, a GPU is a processor that is a specialized electronic circuit designed to rapidly process mathematically intensive applications on electronic devices. The GPU may have a highly parallel structure that is efficient for 40 parallel processing of large blocks of data, such as mathematically intensive data. Alternatively or in addition, one or more of processors Y02 may be a special parallel processing without the graphics optimization, such parallel processors performing the mathematically intensive functions 45 described herein. One or more of processors Y02 may include a processing accelerator (e.g., DSP or other specialpurpose processor).

Computer system Y00 also includes a main memory Y30, and may also include a secondary memory Y40. Main 50 memory Y30 may be a volatile memory or non-volatile memory, and divided into channels. Secondary memory Y40 may include, for example, non-volatile memory such as a hard disk drive Y50, a removable storage drive Y60, and/or a memory stick. Removable storage drive Y60 may com- 55 prise a floppy disk drive, a magnetic tape drive, an optical disk drive, a flash memory, or the like. The removable storage drive Y60 reads from and/or writes to a removable storage unit 470 in a well-known manner. Removable storage unit Y70 may comprise a floppy disk, magnetic tape, 60 optical disk, etc. which is read by and written to by removable storage drive Y60. As will be appreciated by persons skilled in the relevant art(s), removable storage unit Y70 includes a computer usable storage medium having stored therein computer software and/or data.

In alternative implementations, secondary memory Y40 may include other similar means for allowing computer

150

programs or other instructions to be loaded into computer system Y00. Such means may include, for example, a removable storage unit Y70 and an interface (not shown). Examples of such means may include a program cartridge and cartridge interface (such as that found in video game devices), a removable memory chip (such as an EPROM, or PROM) and associated socket, and other removable storage units Y70 and interfaces which allow software and data to be transferred from the removable storage unit Y70 to computer system Y00.

Computer system Y00 may also include a memory controller Y75. Memory controller Y75 controls data access to main memory Y30 and secondary memory Y40. In some embodiments, memory controller Y75 may be external to processor Y10, as shown in FIG. 70. In other embodiments, memory controller Y75 may also be directly part of processor Y10. For example, many AMD<sup>TM</sup> and Intel<sup>TM</sup> processors use integrated memory controllers that are part of the same chip as processor Y10 (not shown in FIG. 70).

Computer system Y00 may also include a communications and network interface Y80. Communication and network interface Y80 allows software and data to be transferred between computer system Y00 and external devices. Communications and network interface Y80 may include a modem, a communications port, a PCMCIA slot and card, or the like. Software and data transferred via communications and network interface Y80 are in the form of signals which may be electronic, electromagnetic, optical, or other signals capable of being received by communication and network interface Y80. These signals are provided to communication and network interface Y80 via a communication path Y85. Communication path Y85 carries signals and may be implemented using wire or cable, fiber optics, a phone line, a cellular phone link, an RF link or other communications channels.

The communication and network interface Y80 allows the computer system Y00 to communicate over communication networks or mediums such as LANs, WANs the Internet, etc. The communication and network interface Y80 may interface with remote sites or networks via wired or wireless connections.

In this document, the terms "computer program medium," "computer-usable medium" and "non-transitory medium" are used to generally refer to tangible media such as removable storage unit Y70, removable storage drive Y60, and a hard disk installed in hard disk drive Y50. Signals carried over communication path Y85 can also embody the logic described herein. Computer program medium and computer usable medium can also refer to memories, such as main memory Y30 and secondary memory Y40, which can be memory semiconductors (e.g. DRAMs, etc.). These computer program products are means for providing software to computer system Y00.

Computer programs (also called computer control logic) are stored in main memory Y30 and/or secondary memory Y40. Computer programs may also be received via communication and network interface Y80. Such computer programs, when executed, enable computer system Y00 to implement embodiments as discussed herein. In particular, the computer programs, when executed, enable processor Y10 to implement the disclosed processes. Accordingly, such computer programs represent controllers of the computer system Y00. Where the embodiments are implemented using software, the software may be stored in a computer program product and loaded into computer system Y00

## US 11,530,454 B2

15

using removable storage drive Y60, interfaces, hard drive Y50 or communication and network interface Y80, for example.

The computer system Y00 may also include input/output/display devices Y90, such as keyboards, monitors, pointing devices, touchscreens, etc.

It should be noted that the simulation, synthesis and/or manufacture of various embodiments may be accomplished, in part, through the use of computer readable code, including general programming languages (such as C or C++), hardware description languages (HDL) such as, for example, Verilog HDL, VHDL, Altera HDL (AHDL), or other available programming tools. This computer readable code can be disposed in any known computer-usable medium including a semiconductor, magnetic disk, optical disk (such as 15 CD-ROM, DVD-ROM). As such, the code can be transmitted over communication networks including the Internet.

The embodiments are also directed to computer program products comprising software stored on any computer-usable medium. Such software, when executed in one or more 20 data processing devices, causes a data processing device(s) to operate as described herein. Embodiments employ any computer-usable or -readable medium, and any computerusable or -readable storage medium known now or in the future. Examples of computer-usable or computer-readable 25 mediums include, but are not limited to, primary storage devices (e.g., any type of random access memory), secondary storage devices (e.g., hard drives, floppy disks, CD ROMS, ZIP disks, tapes, magnetic storage devices, optical storage devices, MEMS, nano-technological storage 30 devices, etc.), and communication mediums (e.g., wired and wireless communications networks, local area networks, wide area networks, intranets, etc.). Computer-usable or computer-readable mediums can include any form of transitory (which include signals) or non-transitory media 35 (which exclude signals). Non-transitory media comprise, by way of non-limiting example, the aforementioned physical storage devices (e.g., primary and secondary storage devices)

It will be understood that any of the embodiments disclosed herein can be used in combination with any other embodiment disclosed herein.

Experimental Section

The presently disclosed embodiments are described in the following Examples, which are set forth to aid in the 45 understanding of the disclosure, and should not be construed to limit in any way the scope of the disclosure as defined in the claims which follow thereafter. The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to use 50 the described embodiments, and is not intended to limit the scope of the disclosure nor is it intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, etc.) 55 but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by volume, and temperature is in degrees Centigrade. It should be understood that variations in the methods as described may be made without changing the fundamental aspects that 60 the experiments are meant to illustrate.

#### Example 1

Exemplary sample preparation and amplification methods 65 are described in U.S. application Ser. No. 13/683,604, filed Nov. 21, 2012; U.S. Publication No. 2013/0123120, and

152

U.S. Ser. No. 61/994,791, filed May 16, 2014, which is hereby incorporated by reference in its entirety. These methods can be used for analysis of any of the samples disclosed herein.

In one experiment, plasma samples were prepared and amplified using a hemi-nested 19,488-plex protocol. The samples were prepared in the following way: up to 20 mL of blood were centrifuged to isolate the buffy coat and the plasma. The genomic DNA in the blood sample was prepared from the buffy coat. Genomic DNA can also be prepared from a saliva sample. Cell-free DNA in the plasma was isolated using the QIAGEN CIRCULATING NUCLEIC ACID kit and eluted in 50 uL TE buffer according to manufacturer's instructions. Universal ligation adapters were appended to the end of each molecule of 40 uL of purified plasma DNA and libraries were amplified for 9 cycles using adaptor specific primers. Libraries were purified with AGENCOURT AMPURE beads and eluted in 50 ul DNA suspension buffer.

6 ul of the DNA was amplified with 15 cycles of STAR 1 (95° C. for 10 min for initial polymerase activation, then 15 cycles of 96° C. for 30 s; 65° C. for 1 min; 58° C. for 6 min; 60° C. for 8 min; 65° C. for 4 min and 72° C. for 30 s; and a final extension at 72° C. for 2 min) using 7.5 nM primer concentration of 19,488 target-specific tagged reverse primers and one library adaptor specific forward primer at 500 nM.

The hemi-nested PCR protocol involved a second amplification of a dilution of the STAR 1 product for 15 cycles (STAR 2) (95° C. for 10 min for initial polymerase activation, then 15 cycles of 95° C. for 30 s; 65° C. for 1 min; 60° C. for 5 min; 65° C. for 5 min and 72° C. for 30 s; and a final extension at 72° C. for 2 min) using reverse tag concentration of 1000 nM, and a concentration of 20 nM for each of 19,488 target-specific forward primers.

An aliquot of the STAR 2 products was then amplified by standard PCR for 12 cycles with 1 uM of tag-specific forward and barcoded reverse primers to generate barcoded sequencing libraries. An aliquot of each library was mixed with libraries of different barcodes and purified using a spin column

In this way, 19,488 primers were used in the single-well reactions; the primers were designed to target SNPs found on chromosomes 1, 2, 13, 18, 21, X and Y. The amplicons were then sequenced using an ILLUMINA GAIIX sequencer. If desired, the number of sequencing reads can be increased to increase the number of targeted SNPs that are amplified and sequenced.

Relevant genomic DNA samples amplified using a seminested 19,488 outer forward primers and tagged reverse primers at 7.5 nM in the STAR 1. Thermocycling conditions and composition of STAR 2, and the barcoding PCR were the same as for the hemi-nested protocol.

#### Example 2

Exemplary primer selection methods are described in U.S. application Ser. No. 13/683,604, filed Nov. 21, 2012 (U.S. Publication No. 2013/0123120) and U.S. Ser. No. 61/994, 791, filed May 16, 2014, which is hereby incorporated by reference in its entirety). These methods can be used for analysis of any of the samples disclosed herein.

The following experiment illustrates an exemplary method for designing and selecting a library of primers that can be used in any of the multiplexed PCR methods of the invention. The goal is to select primers from an initial library of candidate primers that can be used to simultaneously

## US 11,530,454 B2

153

amplify a large number of target loci (or a subset of target loci) in a single reaction volume. For an initial set of candidate target loci, primers did not have to be designed or selected for each target locus. Preferably, primers are designed and selected for a large portion of the most 5 desirable target loci.

Sten 1

A set of candidate target loci (such as SNPs) were selected based on publicly available information about desired parameters for the target loci, such as frequency of the SNPs within a target population or heterozygosity rate of the SNPs (worldwide web at ncbi.nlm.nih.gov/projects/SNP/; Sherry S T, Ward M H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan. 1; 29(1):308-11, which are each incorporated by reference in its entirety). For each candidate locus, one or more PCR primer pairs were designed using the Primer3 program (the worldwide web at primer3.sourceforge.net; libprimer3 release 2.2.3, which is hereby incorporated by reference in its entirety). If there were no feasible designs for PCR 20 primers for a particular target locus, then that target locus was eliminated from further consideration.

If desired, a "target locus score" (higher score representing higher desirability) can be calculated for most or all of the target loci, such as a target locus score calculated based 25 on a weighted average of various desired parameters for the target loci. The parameters may be assigned different weights based on their importance for the particular application that the primers will be used for. Exemplary parameters include the heterozygosity rate of the target locus, the 30 disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, the disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, the specificity of the candidate primer(s) used to amplify the target locus, the size of the candidate primer(s) used to amply the target locus, and the size of the target amplicon. In some embodiments, the specificity of the candidate primer for the target locus includes the likelihood that the candidate primer will mis-prime by binding and amplifying a locus other than the target locus it was designed 40 to amplify. In some embodiments, one or more or all the candidate primers that mis-prime are removed from the library

Step 2

A thermodynamic interaction score was calculated 45 between each primer and all primers for all other target loci from Step 1 (see, e.g., Allawi, H. T. & SantaLucia, J., Jr. (1998), "Thermodynamics of Internal C-T Mismatches in DNA", Nucleic Acids Res. 26, 2694-2701; Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. (1999), 50 "Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A-A, C-C, G-G, and T-T Mismatches", *Biochemistry* 38, 3468-3477; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest-Neighbor Thermodynamics of Internal A-C Mismatches in DNA: Sequence 55 Dependence and pH Effects", Biochemistry 37, 9435-9444; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA", Biochemistry 37, 2170-2179; and Allawi, H. T. & SantaLucia, J., Jr. (1997), "Thermodynamics and 60 NMR of Internal G-T Mismatches in DNA", Biochemistry 36, 10581-10594; MultiPLX 2.1 (Kaplinski L, Andreson R, Puurand T, Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005 Apr. 15; 21(8):1701-2, which are each hereby incorporated by refer- 65 ence in its entirety). This step resulted in a 2D matrix of interaction scores. The interaction score predicted the like154

lihood of primer-dimers involving the two interacting primers. The score was calculated as follows:

interaction score=max(-deltaG\_2,0.8\*(-deltaG\_1))

where

deltaG\_2=Gibbs energy (energy required to break the dimer) for a dimer that is extensible by PCR on both ends, i.e., the 3' end of each primer anneals to the other primer; and deltaG\_1=Gibbs energy for a dimer that is extensible by PCR on at least one end.

Step 3:

For each target locus, if there was more than one primerpair design, then one design was selected using the following method:

For each primer-pair design for the locus, find the worst-case (highest) interaction score for the two primers in that design and all primers from all designs for all other target loci. Pick the design with the best (lowest) worst-case interaction score.

Step 4

A graph was built such that each node represented one locus and its associated primer-pair design (e.g., a Maximal Clique problem). One edge was created between every pair of nodes. A weight was assigned to each edge equal to the worst-case (highest) interaction score between the primers associated with the two nodes connected by the edge.

Step 5

If desired, for every pair of designs for two different target loci where one of the primers from one design and one of the primers from the other design would anneal to overlapping target regions, an additional edge was added between the nodes for the two design. The weight of these edges was set equal to the highest weight assigned in Step 4. Thus, Step 5 prevents the library from having primers that would anneal to overlapping target regions, and thus interfere with each other during a multiplex PCR reaction.

Step 6

An initial interaction score threshold was calculated as follows:

weight\_threshold=max(edge\_weight)-0.05\*(max (edge\_weight)-min(edge\_weight))

where

max(edge\_weight) is the maximum edge weight in the graph; and

min(edge\_weight) is the minimum edge weight in the graph.

The initial bounds for the threshold were set as follows:

max\_weight\_threshold=max(edge\_weight) min\_weight\_threshold=min(edge\_weight)

Step 7

A new graph was constructed consisting of the same set of nodes as the graph from Step 5, only including edges with weights that exceed weight\_threshold. Thus, step ignores interactions with scores equal to or below weight\_threshold. Step 8

Nodes (and all of the edges connected to the removed nodes) were removed from the graph of Step 7 until there were no edges left. Nodes were removed by applying the following procedure repeatedly:

- 1 Find the node with the highest degree (highest number of edges). If there is more than one then pick one arbitrarily.
- 2 Define the set of nodes consisting of the node picked above and all of the nodes connected to it, but excluding any nodes that have degree less than the node picked above.

## US 11,530,454 B2

154

3 Choose the node from the set that has the lowest target locus score (lower score representing lower desirability) from Step 1. Remove that node from the graph. Step 9

If the number of nodes remaining in the graph satisfies the 5 required number of target loci for the multiplexed PCR pool (within an acceptable tolerance), then the method was continued at Step 10.

If there were too many or too few nodes remaining in the graph, then a binary search was performed to determine 10 what threshold values would result in the desired number of nodes remaining in the graphs. If there were too many nodes in the graph then, the weight\_threshold bounds were adjusted as follows:

 $max\_weight\_threshold = weight\_threshold$ 

Otherwise (if there are two few nodes in the graph), then the weight\_threshold bounds were adjusted as follows:

min\_weight\_threshold=weight\_threshold

Then, the weight\_threshold was adjusted follows:

weight\_threshold=(max\_weight\_threshold+min\_ weight\_threshold)/2

Steps 7-9 were repeated. Step 10

The primer-pair designs associated with the nodes remaining in the graph were selected for the library of primers. This primer library can be used in any of the methods of the invention.

If desired, this method of designing and selecting primers can be performed for primer libraries in which only one primer (instead of a primer pair) is used for amplification of a target locus. In this case, a node presents one primer per target locus (rather than a primer pair).

#### Example 3

If desired, methods of the invention can be tested to evaluate their ability to detect a deletion or duplication of a 40 chromosome or chromosome segment. The following experiment was performed to demonstrate the detection of an overrepresentation of the X chromosome or a segment from the X chromosome inherited from the father compared to the X chromosome or X chromosome segment from the 45 mother. This assay is designed to mimic a deletion or duplication of a chromosome or chromosome segment. Different amounts of DNA from a father (with XY sex chromosomes) were mixed with DNA from a daughter (with XX sex chromosomes) of the father for analysis of the extra 50 amount of X chromosome from the father (FIGS. 19A-19D).

DNA from father and daughter cells lines was extracted and quantified using Qubit. Father cell line AG16782, cAG16782-2-F and daughter cell line AG16777, cAG16777-2-P were used. To determine the father's haplotype for the X chromosome, SNPs were detected that are present on the X chromosome but not on the Y chromosome, so there would be a signal from the father's X chromosome but not Y chromosome. The daughter inherited this haplotype from the father. The haplotype from the other X chromosome in the daughter was inherited from her mother. This haplotype from the mother can be determined by assigning the SNPs in the DNA from the daughter cell line that were not inherited from the father to the haplotype from the mother.

To determine whether an overrepresentation of the X chromosome from the father could be detected, different

156

amounts DNA from the father cell line were mixed with DNA from the daughter cell line. The total DNA input was approximately 75 ng (~25 k copies) of genomic DNA. Approximately 3,456 SNPs were amplified using direct multiplex PCR for X and Y chromosome assays. The amplified products were sequenced using 50 bp single run sequencing with 7 bp barcodes using the Rapid/HT mode. The number of reads was approximately 10K per SNP.

As shown in FIGS. 19A-19D, mosaicism from the father's DNA could be detected. These figures indicate that chromosomes segments or entire chromosomes that are overrepresented can be detected.

All patents, patent applications, and published references cited herein are hereby incorporated by reference in their 15 entirety. While the methods of the present disclosure have been described in connection with the specific embodiments thereof, it will be understood that it is capable of further modification. Furthermore, this application is intended to cover any variations, uses, or adaptations of the methods of 20 the present disclosure, including such departures from the present disclosure as come within known or customary practice in the art to which the methods of the present disclosure pertain, and as fall within the scope of the appended claims. Any of the embodiments of the invention 25 can be performed by analyzing the DNA and/or RNA in a sample. For example, any of the methods disclosed herein for DNA can be readily adapted for RNA, for example, by including a reverse transcription step to convert the RNA into DNA.

#### Example 4

This example describes an exemplary method for noninvasive cell-free tumor DNA-based detection of breast 35 cancer-related copy number variations. Breast cancer screening involves mammography, which results in a high false positive rate and misses some cancers. Analysis of tumor-derived circulating cell-free DNA (ctDNA) for cancer-associated CNVs may allow for earlier, safer, and more accurate screening. A SNP-based massively multiplex PCR (mmPCR) approach was used to screen for CNVs in ctDNA isolated from the plasma of breast cancer patients. The mmPCR assay was designed to target 3,168 SNPs on chromosomes 1, 2, and 22, which often have CNVs in cancer (e.g., 49% of breast cancer samples have a 22q deletion). Six plasma samples from breast cancer patientsone stage IIa, four stage IIb, and one stage IIIb-were analyzed. Each sample had CNVs on one or more of the targeted chromosomes. The assay identified CNVs in all six plasma samples, including in one stage IIb sample that was correctly called at a ctDNA fraction of 0.58% (FIGS. 30, 31B, 32A, 32B, and 33); detection only required 86 heterozygous SNPs. A stage IIa sample was also corrected called at a ctDNA fraction of 4.33% using approximately 636 heterozygous SNPs (FIGS. 29, 31A, and 32A). This demonstrates that focal or whole chromosome arm CNVs, both common in cancer, can be readily detected.

To further evaluate sensitivity, 22 artificial mixtures containing a 3 Mb 22q CNV from a cancer cell line were mixed with DNA from a normal cell line (5:95) to simulate a ctDNA fraction of between 0.43% and 7.35% (FIGS. 28A-28C). The method correctly detected CNVs in 100% of these samples. Thus, artificial cfDNA polynucleotide standards/controls can be made by spiking isolated polynucleotide samples that include fragmented polynucleotide mixtures generated by non-cfDNA sources known to exhibit CNV, such as tumor cell lines, into other DNA samples at con-

## US 11,530,454 B2

#### 157

centrations similar to those observed for cfDNA in vivo, such as between, for example, 0.01% and 20%, 0.1 and 15%, or 0.4 and 10% of DNA in that fluid. These standards/controls can be used as controls for assay design, characterization, development, and/or validation, and as quality control standards during testing, such as cancer testing performed in a CLIA lab and/or as standards included in research use only or diagnostic test kits. Significantly, in numerous cancers—including breast and ovarian—CNVs are more prevalent relative to point mutations. Together, this supports that this SNP-based mmPCR approach offers a cost-effective, non-invasive method for detecting these cancers.

#### Example 5

This example describes an exemplary method for detection of copy number variations in breast cancer samples using SNP-targeted massively multiplexed PCR. Evaluation of CNV in tumor tissues typically involves SNP microarray 20 or aCGH. These methods have high whole-genome resolution, but require large amounts of input material, have high fixed costs, and do not work well on formaldehyde fixedparaffin embedded (FFPE) samples. For this example, 28,000-plex SNP-targeted PCR with next generation  $^{25}$ sequencing (NGS) was used to target 1p, 1q, 2p, 2q, 4p16, 5p15, 7q11, 15q, 17p, 22q11, 22q13 and chromosomes 13, 18, 21 and X for detection of CNVs in breast cancer samples. Accuracy was validated on 96 samples with aneuploidies or microdeletions. Single-molecule sensitivity was 30 established by analyzing single cells. Of 17 breast cancer samples (15 fresh frozen and 2 FFPE tumor tissues, 5 pairs of matched tumor and normal cell lines) analyzed, 16 (including both FFPEs) were observed with full or partial CNVs in one to 15 targets (average: 7.8); evidence of tumor heterogeneity was observed. The three tissues with one CNV all had a 1q duplication, the most frequent cytogenetic abnormality in breast carcinoma. The most frequent regions with CNVs were 1q, 7p, and 22q1. Only one tumor tissue (with 9 CNVs) had a region with LOH; this LOH was also 40 detected in adjacent putatively normal tissue that lacked the other 8 CNVs. By contrast, 5 or more regions with LOH and a high total CNV incidence (average: 12.8) was detected in cell lines. Thus, massively multiplexed PCR offers an economical high-throughput approach to investigate CNVs in a 45 targeted manner, and is applicable to difficult-to-analyze samples, such as FFPE tissues.

#### Example 6

This example illustrates exemplary methods for calculating the limit of detection for any of the methods of the invention. These methods were used to calculate the limit of detection for single nucleotide variants (SNVs) in a tumor biopsy (FIG. 34) and a plasma sample (FIG. 35).

The first method (denoted "LOD-mr5" in FIGS. **34** and **35**) calculates the limit of detection based on a minimum of 5 reads being chosen as the minimum number of times a SNV is observed in the sequencing data to have sufficient confidence the SNV is actually present. The limit of detection is based on whether the observed the depth of read (DOR) is above this minimum of 5. The gray lines in FIGS. **34** and **35** indicate SNVs for which the limit of detection is limited by the DOR. In these cases, not enough reads were measured to reach the error limit of the assay. If desired, the 65 limit of detection can be improved (resulting in a lower numerical value) for these SNVs by increasing the DOR.

#### 158

The second method (denoted "LOD-zs5.0" in FIGS. 34 and 35) calculates the limit of detection based on the z-score. The Z-score is the number of standard deviations an observed error percentage is away from the background mean error. If desired, outliers can be removed and the z-score can be recalculated and this process can be repeated. The final weighted mean and the standard deviation of the error rate are used to calculate the z-score. The mean is weighted by the DOR since the accuracy is higher when the DOR is higher.

For the exemplary z-score calculation used for this example, the background mean error and standard deviation were calculated from all the other samples of the same sequencing run weighted by their depth of read, for each genomic locus and substitution type. Samples were not considered in the background distribution if they were 5 standard deviations away from the background mean. The orange lines in FIGS. 34 and 35 indicate SNVs for which the limit of detection is limited by the error rate. For these SNV's enough reads were taken to reach the 5 read minimum, and the limit of detection was limited by the error rate. If desired, the limit of detection can be improved by optimizing the assay to reduce the error rate.

The third method (denoted "LOD-zs5.0-mr5" in FIGS. **34** and **35**) calculates the limit of detection based on the maximum value of the above two metrics.

For the analysis of a tumor sample shown in FIG. **34**, the mean limit of detection was 0.36%, and the median limit of detection was 0.28%. The number of DOR limited (gray lines) SNVs was 934. The number of error rate limited (orange lines) SNVs was 738.

For the analysis of cDNA in a plasma sample shown in FIG. 35, the mean limit of detection was 0.24%, and the median limit of detection was 0.09%. The number of DOR limited (gray lines) SNVs was 732. The number of error rate limited (orange lines) SNVs was 921.

#### Example 7

This example illustrates the detection of CNVs and SNVs from the same single cell. The following primer libraries were used: a library of ~28,000 primers for detecting CNVs, a library of ~3,000 primers for detecting CNVs, and library of primers for detecting SNVs. For analysis of a single cell, cells were serial diluted until there were 3 or 4 cells per droplet. An individual cell was pipetted and placed into a PCR tube. The cell was lysed using Protease K, salt, and DTT using the following conditions: 56° C. for 20 minutes, 95° C. for 10 minutes, and then a 4° C. hold. For analysis of genomic DNA, DNA from the same cell line as the analyzed single cell was either purchased or obtained by growing the cells and extracting the DNA.

For amplification with the library of ~28,000 primers, the following PCR conditions were used: a 40 uL reaction volume, 7.5 nM of each primer, and 2× master mix (MM). In some embodiments QIAGEN Multiplex PCR Kit is used for the master mix (QIAGEN catalog No. 206143; see, e.g., information available at the world wide web at qiagen. com/products/catalog/assay-technologies/end-point-per-and-rt-per-reagents/qiagen-multiplex-per-kit, is which is hereby incorporated by reference in its entirety). The kit includes 2×QIAGEN Multiplex PCR Master Mix (providing a final concentration of 3 mM MgCl<sub>2</sub>, 3×0.85 ml), 5×Q-Solution (1×2.0 ml), and RNase-Free Water (2×1.7 ml). The QIAGEN Multiplex PCR Master Mix (MM) contains a combination of KCl and (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as well as the PCR additive, Factor MP, which increases the local concentration

## US 11,530,454 B2

159

of primers at the template. Factor MP stabilizes specifically bound primers, allowing efficient primer extension by, e.g., HotStarTaq DNA Polymerase. HotStarTaq DNA Polymerase is a modified form of Taq DNA polymerase and has no polymerase activity at ambient temperatures. The following 5 thermocycling conditions were used for the first round of PCR: 95° C. for 10 minutes; 25 cycles of 96° C. for 30 seconds, 65° C. for 29 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and a 4° C. hold. For the second round of PCR a 10 ul reaction volume, 1×MM, and 10 5 nM of each primer was used. The following thermocycling conditions were used: 95° C. for 15 minutes; 25 cycles of 94° C. for 30 seconds, 65° C. for 1 minute, 60° C. for 5 minutes, 65° C. for 5 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and a 4° C. hold.

For the library of ~3,000 primers, exemplary reaction conditions include a 10 ul reaction volume, 2×MM, 70 mM TMAC, and 2 nM primer of each primer. For the library of primers for detecting SNVs, exemplary reaction conditions include a 10 ul reaction volume, 2×MM, 4 mM EDTA, and 20 7.5 nM primer of each primer. Exemplary thermocycling conditions include 95° C. for 15 minutes, 20 cycles of 94° C. for 30 seconds, 65° C. for 15 minutes, and 72° C. for 30 seconds; and then 72° C. for 2 minutes, and a 4° C. hold.

The amplified products were barcoded. One run of <sup>25</sup> sequencing was performed with an approximately equal number of reads per sample.

FIGS. **36**A and **36**B show results from analysis of genomic DNA (FIG. **36**A) or DNA from a single cell (FIG. **36**B) using a library of approximately 28,000 primers <sup>30</sup> designed to detect CNVs. Approximately 4 million reads were measured per sample. The presence of two central bands instead of one central band indicates the presence of a CNV. For three samples of DNA from a single cell, the percent of mapped reads was 89.9%, 94.0%, and 93.4%, <sup>35</sup> respectively. For two samples of genomic DNA the percent of mapped reads was 99.1% for each sample.

FIGS. 37A and 37B show results from analysis of genomic DNA (FIG. 37A) or DNA from a single cell (FIG. 37B) using a library of approximately 3,000 primers 40 designed to detect CNVs. Approximately 1.2 million reads were measured per sample. The presence of two central bands instead of one central band indicates the presence of a CNV. For three samples of DNA from a single cell, the percent of mapped reads was 98.2%, 98.2%, and 97.9%, 45 respectively. For two samples of genomic DNA the percent of mapped reads was 98.8% for each sample. FIG. 38 illustrates the uniformity in DOR for these ~3,000 loci.

For calling SNVs, the call percent for true positive mutations was similar for DNA from a single cell and 50 genomic DNA. A graph of call percent for true positive mutations for single cells on the y-axis versus that for genomic DNA on the x-axis yielded a curve fit of y=1.0076x-0.3088 with R²=0.9834. FIG. 39 shows similar error call metrics for genomic DNA and DNA from a single 55 cell. FIG. 40 shows that the error rate for detecting transition mutations was greater than for detecting transversion mutations, indicating it may be desirable to select transversion mutations for detection rather than transition mutations when possible.

#### Example 8

This example further validates a massively multiplexed PCR methodology for chromosomal aneuploidy and CNV 65 determination disclosed herein, called CoNVERGe (Copy Number Variant Events Revealed Genotypically), and fur-

160

ther illustrates the development and use of "PlasmArt" standards for PCR of ctDNA samples. PlasmArt standards include polynucleotides having sequence identity to regions of the genome known to exhibit CNV and a size distribution that reflects that of cfDNA fragments naturally found in plasma.

Sample Collection

Human breast cancer cell lines (HCC38, HCC1143, HCC1395, HCC1937, HCC1954, and HCC2218) and matched normal cell lines (HCC38BL, HCC1143BL, HCC1395BL, HCC1937BL, HCC1954BL, and HCC2218BL) were obtained from the American Type Culture Collection (ATCC). Trisomy 21 B-lymphocyte (AG16777) and paired father/child DiGeorge Syndrome (DGS) cell lines (GM10383 and GM10382, respectively) were from the Coriell Cell Repository (Camden, N.J.). GM10382 cells only have the paternal 22q11.2 region.

We procured tumour tissues from 16 breast cancer patients, including 11 fresh frozen (FF) samples from Geneticist (Glendale, Calif.) and five formalin-fixed paraffin-embedded (FFPE) samples from North Shore-LIJ (Manhasset, N.Y.). We acquired matched buffy coat samples for eight patients and matched plasma samples for nine patients. FF tumour tissues and matched buffy coat and plasma samples from five ovarian cancer patients were from North Shore-LIJ. For eight breast tumour FF samples, tissue subsections were resected for analysis. Institutional review board approvals from Northshore/LIJ IRB and Kharkiv National Medical University Ethics Committee were obtained for sample collection and informed consent was obtained from all subjects.

Blood samples were collected into EDTA tubes. Circulating tumour DNA was isolated from 1 mL plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Valencia, Calif.).

To make the PlasmArt standards according to one exemplary method, first, 9 □106 cells were lysed with hypotonic lysis buffer (20 mM Tris-Cl (pH 7.5), 10 mM NaCl, and 3 mM MgCl2) for 15 min on ice. Then, 10% IGEPAL CA-630 (Sigma, St. Louis, Mo.) was added to a final concentration of 0.5%. After centrifugation at 3,000 g for 10 min at 4° C., pelleted nuclei were resuspended in 1× micrococcal nuclease (MNase) Buffer (New England BioLabs, Ipswich, Mass.) before adding 1000 U of MNase (New England BioLabs), and then incubated for 5 min at 37° C. Reactions were stopped by adding EDTA to a final concentration of 15 mM. Undigested chromatin was removed by centrifugation at 2,000 g for 1 min. Fragmented DNA was purified with the DNA Clean & Concentrator<sup>TM</sup>-500 kit (Zymo Research, Irvine, Calif.). Mononucleosomal DNA produced by MNase digestion was also purified and size-selected using AMPure XP magnetic beads (Beckman Coulter, Brea, Calif.). DNA fragments were sized and quantified with a Bioanalyzer DNA 1000 chip (Agilent, Santa Clara, Calif.).

To model ctDNA at different concentrations, different fractions of PlasmArts from HCC1954 and HCC2218 cancer cells were mixed with those from the corresponding matched normal cell line (HCC1954BL and HCC2218BL, 60 respectively). Three samples at each concentration were analyzed. Similarly, to model allelic imbalances in plasma DNA in a focal 3.5 Mb region, we generated PlasmArts from DNA mixtures containing different ratios of DNA from a child with a maternal 22q11.2 deletion and DNA from the father. Samples containing only the father's DNA were used as negative controls. Eight samples at each concentration were analyzed.

## US 11,530,454 B2

16

Accordingly, to evaluate the sensitivity and reproducibility of CoNVERGe, especially when the proportion of abnormal DNA for a CNV, or average allelic imbalance (AAI), is low, we used it to detect CNVs in DNA mixtures comprised of a previously characterized abnormal sample titrated into 5 a matched normal sample. The mixtures consisted of artificial cfDNA, termed "PlasmArt", with fragment size distribution approximating natural cfDNA (see above). FIG. 42 graphically displays the size distribution of an exemplary PlasmArt prepared from a cancer cell line compared to the 10 size distribution of cfDNA, looking at CNVs on chromosome arms 1p, 1q, 2p, and 2q. In the first pair, a son's tumor DNA sample having a 3 Mb Focal CNV deletion of the 22q11.2 region was titrated into a matched normal sample from the father at between 0-1.5% total cfDNA (FIG. 41a). CoNVERGe reproducibly identified CNVs corresponding to the known abnormality with estimated AAI of >0.35% in mixtures of ≥0.5%+/-0.2% AAI, failed to detect the CNV in 6/8 replicates at 0.25% abnormal DNA, and reported a value of ≤0.05% for all eight negative control samples. The AAI 20 values estimated by CoNVERGe showed high linearity (R2=0.940) and reproducibility (error variance=0.087). The assay was sensitive to different levels of amplification within the same sample. Based on these data a conservative detection threshold of 0.45% AAI could be used for subsequent 25 analyses. Using this cutoff another experiment was performed in which Plasmart synthetic ctDNA was spiked at known concentrations to create synthetic cancer plasma at between around 0.5% and around 3.5%. Negative plasma was also included as a control. All of the synthetic cancer 30 plasma yielded estimates above 0.45% and the reading for the negative plasma was well below 0.45% (FIG. 43A-D). FIG. 43A shows the maximum likelihood of tumor, and FIG. 43B shows an estimate of DNA fraction results as an odds ratio plot. FIG. 43C is a plot for the detection of transversion 3: events. FIG. 43D is a plot for the detection of Transition

Two additional PlasmArt titrations, prepared from pairs of matched tumor and normal cell line samples and having CNVs on chromosome 1 or chromosome 2, were also 40 evaluated (FIG. 41b, 41c). Among negative controls, all values were <0.45%, and high linearity (R2=0.952 for HCC1954 1p, R2=0.993 for HCC1954 1q, R2=0.977 for HCC2218 2p, R2=0.967 for HCC2218 2q) and reproducibility (error variance=0.190 for HCC1954 1p, 0.029 for 45 HCC1954 1q, 0.250 for HCC2218 2p, and 0.350 for HCC2218 2q) were observed between the known input DNA amount and that calculated by CoNVERGe. The difference in the slopes of the regressions for regions 1p and 1q of one sample pair correlates with the relative difference 50 in copy number observed in the B-allelic frequencies (BAFs) of regions 1p and 1q of the same sample, demonstrating the relative precision of the AAI estimate calculated by CoNVERGe (FIG. 41c, 41d).

The workflow for processing samples is illustrated in FIG. 55 63. CoNVERge has application to a variety of sample sources including FFPE, Fresh Frozen, Single Cell, Germline control and cfDNA. We applied CoNVERGe to six human breast cancer cell lines and matched normal cell lines to assess whether it can detect somatic CNVs. Arm-level and focal CNVs were present in all six tumour cell lines, but were absent from their matched normal cell lines, with the exception of chromosome 2 in HCC1143 in which the normal cell line exhibits a deviation from the 1:1 homolog ratio (FIG. 63b). To validate these results on a different 65 platform, we performed CytoSNP-12 microarray analyses, which produced consistent results for all samples (FIG. 63d,

162

**63***e*). Moreover, the maximum homolog ratios for CNVs identified by CoNVERGe and CytoSNP-12 microarrays exhibited a strong linear correlation (R2=0.987, P<0.001) (FIG. **63***f*).

We next applied CoNVERGe to fresh-frozen (FF) (FIG. 64a) and formalin-fixed, paraffin-embedded (FFPE) breast tumour tissue samples (FIG. 64b, 64d). In both sample types, several arm-level and focal CNVs were present; however, no CNVs were detected in DNA from matched buffy coat samples. CoNVERGe results were highly correlated with those from microarray analyses of the same samples (FIG. 64e-h; R2=0.909, P<0.001 for CytoSNP-12 on FF; R2=0.992, P<0.001 for OncoScan on FFPE). CoNVERGe also produces consistent results on small quantities of DNA extracted from laser capture microdissection (LCM) samples, for which microarray methods are not suitable. Detection of CNVs in Single Cells with CoNVERGe

To test the limits of the applicability of this mmPCR approach, we isolated single cells from the six aforementioned cancer cell lines and from a B-lymphocyte cell line that had no CNVs in the target regions. The CNV profiles from these single-cell experiments were consistent between three replicates and with those from genomic DNA (gDNA) extracted from a bulk sample of about 20,000 cells (FIG. 65). On the basis of the number of SNPs with no sequencing reads, the average assay drop-out rate for bulk samples was 0.48% (range: 0.41-0.60%), which is attributable to either synthesis or assay design failure. For single cells, the additional average assay drop-out rate observed was 0.39% (range: 0.19-0.67%). For single cell assays that did not fail (i.e. no assay drop-out occurred), the average single ADO rate calculated using heterozygous SNPs only was 0.05% (range: 0.00-0.43%). Additionally, the percentage of SNPs with high confidence genotypes (i.e. SNP genotypes determined with at least 98% confidence) was similar for both single cell and bulk samples and the genotype in the single cell samples matched those in the bulk sample (average 99.52%, range: 92.63-100.00%).

In single cells, allele frequencies are expected to directly reflect chromosome copy numbers, unlike in tumour samples where this may be confounded by TH and non-tumour cell contamination. BAFs of 1/n and (n-1)/n indicate n chromosome copies in a region. Chromosome copy numbers are indicated on the allele frequency plots for both single cells and matched gDNA samples (FIG. 65). Application of CoNVERGe to Plasma Samples

To investigate the ability of CoNVERGe to detect CNVs in real plasma samples, we applied our approach to cfDNA paired with a matched tumour biopsy from each of two stage II breast cancer patients and five late-stage ovarian cancer. In all seven patients, CNVs were detected in both FF tumour tissues and in the corresponding plasma samples (FIG. 66). FIG. 67 provides a list of SNV breast cancer mutations. A total of 32 CNVs, at a level of ≥0.45% AAI, were detected in the seven plasma samples (range: 0.48-12.99% AAI) over the five regions assayed, which represent about 20% of the genome. Note that the presence of CNVs in plasma cannot be confirmed due to the lack of alternative orthogonal methods.

Although AAI estimates may appear correlated with BAFs in tumour, direct proportionality should not necessarily be expected due to tumour heterogeneity. For example, in sample BC5 (FIG. 66a), the ovals at the upper left area of FIG. 66a indicate regions that have BAFs compatible with N=11; combining this with the AAI calculation from the plasma sample leads to estimates for c of 2.33% and 2.67% for the two regions. Estimating c using the other regions in

## US 11,530,454 B2

163

the sample give values between 4.46% and 9.53%, which clearly demonstrates the presence of tumor heterogeneity.

These data demonstrate that CNVs can be detected in plasma in a substantial fraction of samples, and suggest that the more prevalent a CNV is within a tumour, the more likely it is to be observed in cfDNA. Furthermore, CoNVERGe detected CNVs from a liquid biopsy that may have otherwise gone unobserved in a traditional tumour biopsy.

#### Example 9

This example provides details regarding certain exemplary sample preparation methods used for CoNVERGe analysis of different types of samples.

Single Cell CNV Protocol for 28,000-Plex PCR

Multiplexed PCR allows simultaneous amplification of many targets in a single reaction. Target SNPs were identified in each genomic region with 10% minimum population minor allele frequency (1000 Genomes Project data; Apr. 30, 2012 release). For each SNP, multiple primers, semi-nested, 20 were designed to have an amplicon length of a maximum length of 75 bp and a melting temperature between 54-60.5° C. Primer interaction scores for all possible combinations of primers were calculated; primers with high scores were eliminated to reduce the likelihood of primer dimer product 25 formation. Candidate PCR assays were ranked and selected on the basis of target SNP minor allele frequency, observed heterozygosity rate (from dbSNP), presence in HapMap, and amplicon length.

In certain experiments, single cell samples were prepared 30 and amplified using a mmPCR 28,000-plex protocol. The samples were prepared in the following way: For analysis of a single cell, cells were serial diluted until there were 3 or 4 cells per droplet. An individual cell was pipetted and placed into a PCR tube. The cell was lysed using Protease K, 35 salt, and DTT using the following conditions: 56° C. for 20 minutes, 95° C. for 10 minutes, and then a 4° C. hold. For analysis of genomic DNA, DNA from the same cell line as the analyzed single cell was either purchased or obtained by growing the cells and extracting the DNA. The DNA was 40 amplified in a 40 uL reaction volume containing Qiagen mp-PCR master mix (2×MM final conc), 7.5 nM primer conc. for 28K primer pairs having a hemi-nested Rev primers under the following conditions: 95 C 10 min, 25× [96 C 30 sec, 65 C 29 min, 72 C 30 sec], 72 C 2 min, 4 C 45 hold. The amplification product was diluted 1:200 in water and 2 ul added to STAR 2 (10 ul reaction volume) 1×MM, 5 nM primer conc. and PCR was performed using heminested inner Fwd primer and tag specific Rev primer: 95 C 15 min, 25× [94 C 30 sec, 65 C 1 min, 60 C 5 min, 65 C 5 50 min, 72 C 30 sec], 72 C 2 min, 4 C hold.

Full sequence tags and barcodes were attached to the amplification products and amplified for 9 cycles using adaptor specific primers. Prior to sequencing, the barcoded library product were pooled, purified with the QIAquick 55 PCR Purification Kit (Qiagen), and quantified using the Qubit® dsDNA BR Assay Kit (Life Technologies). Amplicons were sequenced using an Illumina HiSeq 2500 sequencer.

Extraction of DNA from a Blood/Plasma Sample

Blood samples were collected into EDTA tubes. The whole blood sample was centrifuged and separated into three layers: the upper layer, 55% of the blood sample, was plasma and contains cell-free DNA (cfDNA); the buffy coat middle layer contained leucocytes having DNA, <1% of 65 total; and the bottom layer, 45% of the collected blood sample, contained erythrocytes, no DNA was present in this

164

fraction as erythrocytes are enucleated. Circulating tumor DNA was isolated from at least 1 mL plasma using the QIAamp Circulating Nucleic Acid Kit, Qia-Amp (Qiagen, Valencia, Calif.), according to the manufacture's protocol. Plasma CNV Protocol for 3,168-plex for Chromosomes 1p, 1q, 2p, 2q, and 22q11

Plasma DNA libraries were prepared and amplified using a mmPCR 3,168-plex protocol. The samples were prepared in the following way: Up to 20 mL of blood was centrifuged to isolate the buffy coat and the plasma. Plasma extraction of cfDNA and library preparation was performed. DNA was eluted in 50 uL TE buffer. The input for mmPCR was 6.7~uLof amplified and purified Natera plasma library at an input amount of approximately 1200 ng. The plasma DNA was amplified in a 20 uL reaction volume containing Qiagen mp-PCR master mix (2×MM final conc), 2 nM tagged primer conc. (total 12.7 uM) in 3,168-plex primer pools and PCR amplified: 95 C 10 min, 25× [96 C 30 sec, 65 C 20 min, 72 C 30 sec], 72 C 2 min, 4 C hold. The amplification product was diluted 1:2,000 in water and 1 ul added to the Barcoding-PCR in a 10 uL reaction volume. The barcodes are attached to the amplification products via PCR amplification for 12 cycles using tag specific primers. Products of multiple samples are pooled and then purified with QIAquick PCR Purification Kit (Qiagen) and eluted in 50 ul DNA suspension buffer. Samples are sequenced by NGS as described for the Single Cell CNV Protocol for 28,000-plex PCR.

Breast Cancer Feasibility SNV Panel from Plasma

cfDNA from breast cancer patient blood samples was prepared and amplified using 336 primer pairs that were distributed into four 84-plex pools. Natera plasma libraries were prepared as described for Plasma CNV Protocol for 3,168-plex for Chromosomes 1p, 1q, 2p, 2q, and 22q11. DNA was eluted in 50 uL TE buffer. The input for mPCR was 2.5 uL of amplified and purified Natera plasma library at an input amount of approximately 600 ng. FIG. 68A-B represents the major and minor allele frequencies of the SNPs used in a 3168 mmPCR reaction. The X-axis represents the number of SNPs, from left to right, for chromosome 1q, 1p, 2q, 2p and 22q. SNPs were selected from the 1000 Genomes map for Humans, Group 19 and dbSNP to pick targets, but only SNPs from the 1000 Genomes were used to screen for minor allele frequencies. The plasma DNA was amplified in four parallel reactions of 84-plex primer pools, a 10 uL reaction volume containing Qiagen mp-PCR master mix (2×MM final conc.), 4 mM EDTA, 7.5 nM primer concentration (total 1.26 uM) and PCR amplified: 95 C 15 min, 25× [94 C 30 sec, 65 C 15 min, 72 C 30 sec], 72 C 2 min, 4 C hold. The amplification product of the 4 subpools were each diluted 1:200 in water and 1 ul added to the Barcoding-PCR reaction in a 10 uL reaction volume containing Q5 HS HF master mix (1×final), and 1 uM each barcoding primer and each of the pools were amplified in the following reaction: 98 C 1 min, 25× [98 C 10 sec, 70 C 10 sec, 60 C 30 sec, 65 C 15 sec, 72 C 15 sec], 72 C 2 min, 4 C hold. Libraries were purified with QIAquick PCR Purification Kit (Qiagen) and eluted in 50 ul DNA suspension buffer. Samples were sequenced by paired end sequencing.

# Example 10

This example provides details regarding certain exemplary methods for analyzing sequencing data to identify SNVs.

SNV Method 1:

For this embodiment, a background error model was constructed using normal plasma samples, which were sequenced on the same sequencing run to account for

run-specific artifacts. In certain embodiments, 5, 10, 15, 20, 25, 30, 40, 50, 100, 150, 200, 250, or more than 250 normal plasma samples were analyzed on the same sequencing run. In certain illustrative embodiments, 20, 25, 40, or 50 normal plasma samples are analyzed on the same sequencing run. 5 Noisy positions with normal median variant allele frequency greater than a cutoff are removed. For example this cutoff in certain embodiments is >0.1%, 0.2%, 0.25%, 0.5%, 1%, 2%, 5%, or 10%. In certain illustrative embodiments noisy positions with normal medial variant allele frequency 10 greater than 0.5% are removed. Outlier samples were iteratively removed from the model to account for noise and contamination. In certain embodiments, samples with a Z score of greater than 5, 6, 7, 8, 9, or 10 were removed from the data analysis. For each base substitution of every genomic loci, the depth of read weighted mean and standard deviation of the error were calculated. Tumor or cell-free plasma samples' positions with at least 5 variant reads and a Z-score of 10 against the background error model were called as a candidate mutation.

## SNV Method 2:

For this embodiment we aim to determine Single Nucleotide Variants (SNVs) using plasma ctDNA data. We model the PCR process as a stochastic process, estimate the parameters using a training set and make the final SNV calls using a separate testing set. The main idea is to determine the propagation of the error across multiple PCR cycles, calculate the mean and the variance of the background error, and differentiate the background error from real mutations.

The following parameters are estimated for each base: p=efficiency (probability that each read is replicated in each cycle)

 $p_e$ =error rate per cycle for mutation type e (probability that an error of type e occurs)

X<sub>0</sub>=initial number of molecules

As a read is replicated over the course of PCR process, the more errors occur. Hence, the error profile of the reads is determined by the degrees of separation from the original read. We refer to a read as *k*<sup>th</sup> generation if it has gone through k replications until it has been generated.

Let us define the following variables for each base:  $X_{ij}$ =number of generation i reads generated in the PCR

cycle j  $Y_{ij}$ =total number of generation i reads at the end of cycle 45

 $X_{ij}^{e}$ =number of generation i reads with mutation e generated in the PCR cycle j

Moreover, in addition to normal molecules  $X_0$ , if there are additional  $f_e X_0$  molecules with the mutation e at the 50 beginning of the PCR process (hence  $f_e/(1+fe)$  will be the fraction of mutated molecules in the initial mix-

Given the total number of generation i–1 reads at cycle j–1, the number of generation i reads generated at cycle 55 j has a binomial distribution with a sample size of  $Y_{i-1,j-1}$  and probability parameter of p. Hence,  $E(X_{ij}, | Y_{i-1,j-1}, p) = p Y_{i-1,j-1}$  and  $Var(X_{ij}, | Y_{i-1,j-1}, p) = p(1-p)$ 

We also have  $Y_{ij} = \Sigma_{k=i} X_{ik}$ . Hence, by recursion, simulation or similar methods, we can determine  $E(X_{ij})$ . Similarly, we can determine  $Var(X_{ij}) = E(Var(X_{ij}, |p)) + Var(E(X_{ij}, |p))$  using the distribution of p.

Var(E( $X_{ij}$ , |p)) using the distribution of p. finally, E( $X_{ij}^e$ |Y<sub>i-1,j-1</sub>, p<sub>e</sub>)=p<sub>e</sub> Y<sub>i-1,j-1</sub> and Var( $X_{ij}^e$ |Y<sub>i-1,j-1</sub>, p)=p<sup>e</sup> (1-p<sub>e</sub>) Y<sub>i-1,j-1</sub>, and we can use these to compute 65 E( $X_{ij}^e$ ) and Var( $X_{ij}^e$ ).

166

6+0.2 Algorithm

The algorithm starts by estimating the efficiency and error rate per cycle using the training set. Let n denote the total number of PCR cycles.

The number of reads Rb at each base b can be approximated by  $(1+p_b)^n X_0$ , where  $p_b$  is the efficiency at base b. Then  $(R_b/X_0)^{1/n}$  can be used to approximate  $1+p_b$ . Then, we can determine the mean and the standard variation of  $p_b$  across all training samples, to estimate the parameters of the probability distribution (such as normal, beta, or similar distributions) for each base.

Similarly, the number of error e reads  $R_b^{\ e}$  at each base b can be used to estimate  $p_e$ . After determining the mean and the standard deviation of the error rate across all training samples, we approximate its probability distribution (such as normal, beta, or similar distributions) whose parameters are estimated using this mean and standard deviation values.

Next, for the testing data, we estimate the initial starting 20 copy at each base as

$$\int_0^1 \frac{R_b}{(1+p_b)^n} f(p_b) dp_b$$

where f(.) is an estimated distribution from the training set.

$$\int_0^1 \frac{R_b}{(1+p_b)^n} f(p_b) dp_b$$

where f(.) is an estimated distribution from the training set.

Hence, we have estimated the parameters that will be used
in the stochastic process. Then, by using these estimates, we can estimate the mean and the variance of the molecules created at each cycle (note that we do this separately for normal molecules, error molecules, and mutation molecules).

Finally, by using a probabilistic method (such as maximum likelihood or similar methods), we can determine the best  $f_e$  value that fits the distribution of the error, mutation, and normal molecules the best. More specifically, we estimate the expected ratio of the error molecules to total molecules for various  $f_e$  values in the final reads, and determine the likelihood of our data for each of these values, and then select the value with the highest likelihood.

In certain embodiments, Method 2 above is performed as follows:

- a) Estimate a PCR efficiency and a per cycle error rate using a training data set;
- b) Estimate a number of starting molecules for the testing data set at each base using the distribution of the efficiency estimated in step (a);
- c) If needed, update the estimate of the efficiency for the testing data set using the starting number of molecules estimated in step (b);
- d) Estimate the mean and variance for the total number of molecules, background error molecules and real mutation molecules (for a search space consisting of an initial percentage of real mutation molecules) using testing set data and parameters estimated in steps (a), (b) and (c);
- e) Fit a distribution to the number of total error molecules (background error and real mutation) in the total molecules, and calculate the likelihood for each real mutation percentage in the search space; and

## US 11,530,454 B2

#### 16

f) Determine the most likely real mutation percentage and calculate the confidence using the data from in step (e).

#### Example 11

This example provides results using the multiplexed PCR CoNVERGe methods provided herein, for the detection of cancer by detecting CNV in circulating DNA. The Plasma CNV Protocol for 3,168-plex for Chromosomes 1p, 1q, 2p, 2q, and 22q11 provided herein, was used. Plasma from 21 breast cancer patients (stage I-IIIB) was analyzed. The results shown in FIG. 44 demonstrate that CNVs were detected in all samples using an AAI>=0.45% and required as few as 62 heterozygous SNPs. A similar protocol was used to analyze plasma from ovarian cancer patients. Using a 0.45% cutoff, a 100% ovarian cancer detection rate was achieved, as shown in FIG. 45. Each of the five samples also had a matched tumor sample.

#### Example 12

This example demonstrates that a dramatic improvement in the ability to detect cancer is achieved by testing plasma for the presence of CNVs and SNVs. CNVs and SNVs were detected using the methods provided in the Examples above. Samples were prepared according to the appropriate protocols in Example 9. SNVs were identified using SNV Method 1 above. As shown in FIG. 46, the sensitivity of detecting breast and lung cancer are greatly improved by analyzing plasma from Stage I-III cancer patients for both CNVs and SNVs versus testing for SNVs alone. Analyzing SNVs only, 71% of cancers were detected in plasma samples. However by analyzing for the presence of SNVs and/or CNVs the detection rate goes up to 83% for breast and 92% for lung in the patient populations analyzed. If one considers all of the SNVs and CNVs that have been identified in the TCGA and COSMIC data sets, the expected diagnostic load would 3 be greater than 97% for breast cancer and >98% for lung

Further analysis was performed on samples from 41 patient samples with different stages of cancer using the plasma sample prep methods provided in Example 9 and 40 SNV Method 1 provided above. As shown in FIG. 47, when assaying for CNVs and SNVs in circulating tumor DNA from breast cancer patients 60% of Stage I, 88% of Stage II and 100% of Stage III breast cancers were detected using a limit of quantification of 0.2% ctDNA for SNVs and 0.45% ctDNA for CNVs. As shown in FIG. 48, when assaying for CNVs and SNVs in ctDNA and looking at 41 patient samples with different substages of breast cancer, 60% of Stage I, 100% of Stage II, 90% of Stage IIA, 80% of Stage IIB, and 100% of Stage III, IIIA, and IIIB breast cancers were detected using a limit of quantification of 0.2% ctDNA for SNVs and 0.45% ctDNA for CNVs. As shown in FIG. 49, when assaying for CNVs and SNVs in 24 circulating tumor DNA from lung cancer patient samples 88% of Stage I, 100% of Stage II and 100% of Stage III lung cancers were detected using a limit of quantification of 0.2% ctDNA for 55 SNVs and 0.45% ctDNA for CNVs. As shown in FIG. 50, when assaying for CNVs and SNVs in ctDNA and looking at 24 patient samples with different substages of lung cancer, 100% detection rate was achieved for all substages except that an 82% detection rate was achieved for the patients with 60 stage IB lung cancer using a limit of quantification of 0.2% ctDNA for SNVs and 0.45% ctDNA for CNVs.

#### Example 13

This example demonstrates that detection of SNV in ctDNA overcomes the limitations in identifying variant

#### 168

alleles in biopsied samples due to tumor heterogeneity. TRACERx samples of three small cell lung cancer patient samples and one adenocarcinoma lung cancer patient sample for which tumor biopsies and corresponding pre-operative blood plasma samples had been collected were used for analysis of tumor heterogeneity. Samples were obtained from the Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London WC1E 6BT, UK. Samples were primary lung cancer samples for analysis of SNV mutations. Two to three biopsies from various regions from the entire cancerous lung were taken from each patient (FIG. 51A). Each biopsied sample was assayed by whole exome sequencing (Illumina HiSeq200; Illumina, San Diego, Calif.), followed by AmpliSeq® sequencing (Ion Torrent, South San Francisco, Calif.) on a PGM® for identification of underlying clonal heterogeneity. Following sequencing and SNV analyses, the variant allele frequency (VAF) was determined for each biopsy sample (FIG. 51B).

Plasma samples from each of the four patients were used to isolate ctDNA and identify both clonal and subclonal SNV mutations in plasma to overcome tumor heterogeneity (FIG. 52). Clonal populations had VAF allele calls in all biopsied samples assayed and in plasma while subclonal populations had VAF allele calls in at least one biopsy sample, but not all biopsy samples. The plasma was considered to be a cumulative representative of the SNV's found in the ctDNA of each patient. Not all SNV's identified by sequencing were able to have corresponding PCR assays designed.

To compare the AmpliSeq (Swanton) and mmPCR/NGS assay methods for identifying tumor heterogeneity, Natera designed PCR assays for each SNV mutation for VAF detection in both biopsied and corresponding ctDNA from plasma (FIG. 53). Blank cells represent no biopsy sample available and a zero value represents no VAF detected. The following 11 genes were initially identified as a negative (false VAF call) by the AmpliSeq FP or FN assays but were called correctly by the Natera TP or TN assays and mmPCR/ NGS assay methods: L12: CYFIP1, FAT1, MLLT4, and RASA1; L13: HERC4, JAK2, MSH2, MTOR, and PLCG2; L15: GABRG1; L17: TRIM67. Surprisingly, when the AmpliSeq raw sequencing data was re-examined these results were verified. The raw AmpliSeq data sequencing files revealed that the data fell below the PGM or Illumina detectable threshold setting. The data identified 16/38 variants were detected in plasma and that there were several biopsy samples in the L12 patient samples that had predominant clonal SNV mutations: L12: BRIP1, CARS, FAT1, MLLT4, NFE2L2, TP53, TP53 as well as patients L13: EGFR, EGFR, TP53 and L15: KDM6A, ROS1. An additional two patients were found to have a total of four subclonal variant mutations in plasma: L12: CIC, KDM6A and L17; NF1, TRIM67. These results are summarized in FIG. 54A which is a whisker plot of the mean VAF for each sample listed in FIG. 53 by each assay method and FIG. 54B is a direct comparison represented by a linear regression plot of each assay's VAF sample mean.

#### Example 14

This example demonstrates that by using low primer concentrations such that primer amount is the limiting reactant in multiplex PCR in a workflow that is followed by next generation sequencing, uniformity of density of reads, and therefore limits of detection, across a pool of amplification reactions is improved. Some experiments were car-

## US 11,530,454 B2

169

ried out for plasma CNV using the 3,168-plex panel according to Example 9 above except that the total reaction volume was 10 uL instead of 20 uL. Furthermore, PCR was carried out for 15, 20, or 25 cycles. Other experiments were carried out using the four 84-plex pools on breast cancer samples according to the protocol of Example 9 except that primer concentrations were 2 nM and PCR amplification was carried out for 15, 20, or 25 cycles.

Not to be limited by theory, it is believed that primer limited multiplex PCR provides improved depth of read 10 uniformity for multiplex PCR before multi-read sequencing, such as sequencing on an Illumina HiSeq or MiSeq system or an Ion Torrent PGM or Proton system, based on the following considerations: If some of the amplifications in a multiplex PCR have lower efficiencies than others, then with 15 normal multiplex PCR we will end up with a wide range of depth of read ("DOR") values. However, if the amount of primer is limited, and the multiplex PCR is cycled more times than what it takes to exhaust the primers, then the more efficient amplifications will stop doubling (because they 20 have no more primers to use) and the less efficient ones will continue to double; this will result in a more similar amount of amplification product for all of the amplification products. This will translate into a much more uniform distribution of the DOR.

The following calculations are used to determine the number of cycles that would exact a given amount of primer and starting nucleic acid template:

assume a given starting DNA input level: 100 k copies of each target (10°5; this is easily achieved with using 30 amplified library)

assume we use 2 nM of each primer as an exemplary concentration, although other concentrations such as, for example, 0.2, 0.5, 1, 1.5, 2, 2.5, 5, or 10 nM could work too.

calculate the number of primer molecules for each primer: 2\*10^-9 (molar concentration, 2 nM)×10\*10^-6 (reaction volume, 10 ul)×6\*10^23 (number of molecules per mole, Avogadro's number)=12\*10^9

calculate the amplification fold needed to consume all 40 primers: 12\*10^9 (number of primer molecules)/10^5 (number of copies of each target)=12\*10^4

calculate the number of cycles needed to achieve this amplification fold, assuming 100% efficiency at each cycle: log 2(12\*10^4)=17 cycles. (this is log 2 because 45 at each cycle, the number of copies doubles).

So for these conditions (100 k copies input, 2 nM primers, 10 ul reaction volume, assuming 100% PCR efficiency at each cycle), the primers would be consumed after 17 PCR cycles

However, the key assumption is that some of the products DO NOT have 100% efficiency, so without measuring their efficiencies (which is only practicable for a small number of them anyway), it would take more than 17 cycles to consume them.

FIGS. **55-58** show results for the four 84-plex SNV PCR primer pools. For each of the pools we observed improved DOR efficiency with increasing cycles from 15 to 20 to 25. Similar results were obtained for experiments using the 3,168-plex panel (FIGS. **59-61**). The limit of detection 60 decreased (i.e. SNV sensitivity increased) with increasing depth of read. Furthermore, the sensitivity was consistently better when detecting transversion mutations than transition mutations. It is likely that additional increases in DOR efficiency can be obtained with additional cycles when using primer-limiting multiplex PCR before multi-read sequenc-

170

Accordingly, in one aspect provided herein is a method of amplifying a plurality of target loci in a nucleic acid sample that includes (i) contacting the nucleic acid sample with a library of primers and other primer extension reaction components to provide a reaction mixture, wherein the relative amount of each primer in the reaction mixture compared to the other primer extension reaction components creates a reaction wherein the primers are present at a limiting concentration, and wherein the primers hybridize to a plurality of different target loci; and (ii) subjecting the reaction mixture to primer extension reaction conditions for sufficient number of cycles to consume or exhaust the primers in the library of primers, to produce amplified products that include target amplicons. For example, the plurality of different target loci can include at least 2, 3, 5, 10, 25, 50, 100, 200, 250, 500, 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; or 100,000 different target loci, and at most, 50, 100, 200, 250, 500, 1,000; 2,000; 5,000; 7,500; 10,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; 100,000, 200,000, 250,000, 500, 000, and 1,000,000 different target loci to produce a reaction mixture.

The method in illustrative embodiments, includes determining an amount of primer that will be a rate limiting 25 amount. This calculation typically includes estimating and/ or determining the number of target molecules and involves analyzing and/or determining the number of amplification cycles performed. For example, in illustrative embodiments, the concentration of each primer is less than 100, 75, 50, 25, 10, 5, 2, 1, 0.5, 0.25, 0.2 or 0.1 nM. In various embodiments, the GC content of the primers is between 30 to 80%, such as between 40 to 70% or 50 to 60%, inclusive. In some embodiments, the range of GC content (e.g., the maximum GC content minus minimum GC content, such as 80%-60%=a range of 20%) of the primers is less than 30, 20, 10, or 5%. In some embodiments, the melting temperature  $(T_m)$ of the primers is between 40 to 80° C., such as 50 to 70° C., 55 to 65° C., or 57 to 60.5° C., inclusive. In some embodiments, the range of melting temperatures of the primers is less than 20, 15, 10, 5, 3, or 1° C. In some embodiments, the length of the primers is between 15 to 100 nucleotides, such as between 15 to 75 nucleotides, 15 to 40 nucleotides, 17 to 35 nucleotides, 18 to 30 nucleotides, 20 to 65 nucleotides, inclusive. In some embodiments, the primers include a tag that is not target specific, such as a tag that forms an internal loop structure. In some embodiments, the tag is between two DNA binding regions. In various embodiments, the primers include a 5' region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In various embodiments, the length of the 3' region is at least 7 nucleotides. In some embodiments, the length of the 3' region is between 7 and 20 nucleotides, such as between 7 to 15 nucleotides, or 7 to 10 nucleotides, inclusive. In various embodiments, the test primers include a 5' region that is not specific for a target locus (such as a tag or a universal primer binding site) followed by a region that is specific for a target locus, an internal region that is not specific for the target locus and forms a loop structure, and a 3' region that is specific for the target locus. In some embodiments, the range of the length of the primers is less than 50, 40, 30, 20, 10, or 5 nucleotides. In some embodiments, the length of the target amplicons is between 50 and 100 nucleotides, such as between 60 and 80 nucleotides, or 60 to 75 nucleotides, inclusive. In some embodiments, the range of the length of the target amplicons is less than 100, 75, 50, 25, 15, 10, or 5 nucleotides.

## US 11,530,454 B2

#### 171

In various embodiments of any of the aspects of the invention, the primer extension reaction conditions are polymerase chain reaction conditions (PCR). In various embodiments, the length of the annealing step is greater than 3, 5, 8, 10, or 15 minutes but less than 240, 120, 60, or 30 minutes. In various embodiments, the length of the extension step is greater than 3, 5, 8, 10, or 15 minutes but less than 240, 120, 60 or 30 minutes.

#### Example 15

This Example demonstrates the ability of the SNV detection methods of the present invention to identify mosaicism in single cell analysis also referred to as single molecule analysis. FIG. **62** shows multiplex PCR results from tumor cell genomic DNA and single cell/molecule inputs using the 28K-plex primer set according to the 28K single cell method provided in Example 9. Using this method, greater than 85% of reads were mapped—over 4.7M reads (about 167 reads per target). The lower portion of the figure shows that mosaicism was observed among cells.

#### What is claimed is:

- 1. A method for preparing a plasma sample of a subject 25 having cancer or suspected of having cancer useful for detecting one or more single nucleotide variant (SNV) mutations in the plasma sample, the method comprising:
  - performing whole exome sequencing or whole genome sequencing on a tumor sample of the subject to identify 30 a plurality of tumor-specific SNV mutations;
  - performing targeted multiplex amplification to amplify 10 to 500 target loci each encompassing a different tumor-specific SNV mutation from cell-free DNA isolated from a plasma sample of the subject or DNA derived 35 therefrom to obtain amplicons having a length of 50-150 bases, wherein the target loci are amplified together in the same reaction volume; and
  - sequencing the amplicons to obtain sequence reads, and detecting one or more of the tumor-specific SNV mutations present in the cell-free DNA from the sequence reads, wherein the sequencing has a depth of read of at least 50,000 per target locus.
- 2. The method of claim 1, wherein the cell-free DNA comprises circulating tumor DNA.
- 3. The method of claim 1, wherein the SNV mutations comprise one or more clonal SNV mutations.
- **4.** The method of claim **1**, wherein the SNV mutations comprise one or more subclonal SNV mutations.
- **5**. The method of claim **1**, wherein the SNV mutations 50 comprise one or more clonal SNV mutations and one or more subclonal SNV mutations.
- **6**. The method of claim **1**, wherein the tumor sample of the subject is a tumor tissue sample.
- 7. The method of claim 1, wherein the method further 55 comprises determining clonal heterogeneity of the tumor sample.
- **8**. The method of claim **1**, wherein the targeted multiplex amplification amplifies 20 to 50 target loci each encompassing a different tumor-specific SNV mutation.
- 9. The method of claim 1, wherein the targeted multiplex amplification amplifies 50 to 100 target loci each encompassing a different tumor-specific SNV mutation.
- 10. The method of claim 1, wherein the method further comprises designing PCR primers or hybrid capture probes 65 targeting the plurality of SNV mutations identified in the tumor sample.

#### 172

- 11. The method of claim 1, wherein the method further comprises performing barcoding PCR prior to the sequencing.
- 12. The method of claim 1, wherein the method further comprises detecting recurrence and/or metastases of the cancer from the SNV mutations detected in the cell-free DNA
- 13. The method of claim 1, wherein the cancer is colorectal cancer, lung cancer, bladder cancer, or breast cancer.
- **14.** A method for preparing a plasma sample of a subject having cancer or suspected of having cancer useful for detecting one or more single nucleotide variant (SNV) mutations in the plasma sample, the method comprising:
  - performing whole exome sequencing or whole genome sequencing on a tumor sample of the subject to identify a plurality of tumor-specific SNV mutations;
  - performing targeted multiplex amplification to amplify 10 to 500 target loci each encompassing a different tumor-specific SNV mutation from cell-free DNA isolated from a plasma sample of the subject or DNA derived therefrom to obtain amplicons, wherein the target loci are amplified together in the same reaction volume; and
  - sequencing the amplicons to obtain sequence reads, and detecting one or more of the tumor-specific SNV mutations present in the cell-free DNA from the sequence reads, wherein the method is capable of detecting an SNV mutation that is present in less than or equal to 0.015% of the cell-free DNA comprising the SNV locus.
- 15. The method of claim 14, wherein the sequencing has a depth of read of at least 50,000 per target locus.
- **16**. The method of claim **14**, wherein the cell-free DNA comprises circulating tumor DNA.
- 17. The method of claim 14, wherein the SNV mutations comprise one or more clonal SNV mutations.
- 18. The method of claim 14, wherein the SNV mutations comprise one or more subclonal SNV mutations.
- 19. The method of claim 14, wherein the SNV mutations comprise one or more clonal SNV mutations and one or more subclonal SNV mutations.
- 20. The method of claim 14, wherein the tumor sample of the subject is a tumor tissue sample.
- 21. The method of claim 14, wherein the method further comprises determining clonal heterogeneity of the tumor sample.
- 22. The method of claim 14, wherein the targeted multiplex amplification amplifies 20 to 50 target loci each encompassing a different tumor-specific SNV mutation.
- 23. The method of claim 14, wherein the targeted multiplex amplification amplifies 50 to 100 target loci each encompassing a different tumor-specific SNV mutation.
- 24. The method of claim 14, wherein the method further comprises designing PCR primers or hybrid capture probes targeting the plurality of SNV mutations identified in the tumor sample.
- 25. The method of claim 14, wherein the method further comprises performing barcoding PCR prior to the sequencing.
- 26. The method of claim 14, wherein the method further comprises detecting recurrence and/or metastases of the cancer from the SNV mutations detected in the cell-free DNA.
- 27. The method of claim 14, wherein the cancer is colorectal cancer, lung cancer, bladder cancer, or breast cancer

US 11,530,454 B2

173

28. The method of claim 14, wherein the method is capable of detecting an SNV mutation that is present in 0.005% to 0.015% of the cell-free DNA comprising the SNV locus.

\* \* \* \*